

Univerzita Karlova v Praze

Přírodovědecká fakulta

Katedra Biochemie



Barbora Ptáčková

Komplementační peptid LacZ-alfa jako nástroj pro studium
molekulární evoluce

LacZ-alpha complementation peptide as a tool for molecular
evolution studies

Bakalářská práce

Vedoucí závěrečné práce: Mgr. Klára Hlouchová, Ph.D.

Praha, 2015

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně pod vedením školitelky Mgr. Kláry Hlouchové, Ph.D. a všechny použité prameny jsem řádně citovala.

V Praze, dne.....

Podpis.....

Poděkování

Ráda bych na tomto místě poděkovala své školitelce Mgr. Kláře Hlouchové, Ph.D. za cenné rady, pomoc, ochotu, trpělivost a vřelý přístup při vypracovávání této práce.

Dále bych ráda poděkovala RNDr. et PharmDr. Taťáně Majerové, Ph.D. za všestrannou pomoc po celou dobu mé práce v laboratoři.

Můj velký dík patří Doc. RNDr. Janu Konvalinkovi, CSc. za příležitost k vypracování bakalářské práce v jeho laboratoři a za konzultace ohledně této práce.

V neposlední řadě děkuji své rodině za bezmeznou podporu během celé doby mého studia.

Abstrakt

Proteiny jsou stavební a funkční podstatou všech živých organismů. Ačkoliv o jejich strukturních a funkčních vlastnostech dnes víme poměrně mnoho, jejich evoluce je pro nás stále nejasná.

V současné době převažuje názor, že prvotní genetický kód obsahoval menší počet aminokyselin a postupem evoluce k němu byly přiřazovány aminokyseliny evolučně novější. Obecně uznávané teorie a studie předpokládají dva zdroje raných aminokyselin (exogenní a endogenní). Z rozsáhlého repertoáru prebioticky dostupných aminokyselin je v současnosti jen 10 geneticky kódovaných, tedy proteinogenních. Zbylých 10 proteinogenních (tedy evolučně novějších) aminokyselin zřejmě vzniklo až v důsledku evoluce biosyntetických drah. Otázkou zůstává, je-li raná polovina současných proteinogenních aminokyselin dostačující pro tvorbu strukturovaných a funkčních proteinů, které mohly dát vznik prvotnímu metabolismu. A je v současném světě možné nahradit evolučně novější aminokyseliny těmi ranými za současného udržení strukturně-funkční integrity? Právě tyto otázky tvoří základ předkládané práce.

Cílem této práce bylo připravit metodiku pro mutagenezi zkráceného LacZ-alfa peptidu za účelem pozdějšího studia randomizace evolučně novějších aminokyselin za rané. Jako modelový protein byl vybrán LacZ-alfa peptid, který je v molekulární biologii využíván při procesu α -komplementace. Jako vhodný nástroj pro počáteční testování knihoven vzniklých variant se ukázala být metoda modro/bílé selekce, která využívá procesu α -komplementace. V rámci redukce studovaného systému bylo ověřeno, že zkrácený pUC19 LacZ-alfa peptid (obsahující zbytky 5 – 41) je pro účely α -komplementace funkční. Dále byla ověřena funkčnost metody mutageneze vybraného zbytku (Trp36), tvorby knihoven možných variant a jejich selekce pro analýzu fenotypu a genotypu.

Klíčová slova: evoluce proteinů, genetický kód, aminokyseliny, mutageneze

Abstract

Proteins are the key structural and functional molecules of living organisms. Although the last decades have brought a lot of knowledge about their structural and functional characteristics, science still lacks very basic answers about how these properties evolved.

Current predominant opinions suggest that early genetic code contained only a subset of today's canonical amino acids. Both exogenous and endogenous sources of prebiotic amino acids imply that even though the prebiotic amino acid repertoire was very broad, only about half of the proteinogenic amino acids were present. It follows that the “evolutionary new” amino acids were added to the genetic coding system only after the evolution of their biosynthetic pathways. From the current scientific knowledge it is unclear whether proteins composed of “evolutionary old” amino acids could serve basic metabolic functions and if today's proteins could be “reversely-evolved” to be composed of only such a subset of amino acids while maintaining their structural and functional integrity. These questions lie at the core of this study.

This thesis aims to test a starting methodology that would randomize “new” amino acid positions by “old” amino acids in the sequence of LacZ-alpha peptide. This peptide was selected as a target model protein because it is used in the process of α -complementation during blue/white screening and is thus a key part of a routine molecular biology technique that could be used for testing initial libraries of generated LacZ-alpha variants. Within this thesis it was confirmed that in order to reduce the LacZ-alpha sequence to a minimal space, an abbreviated version (containing residues 5-41) is functional in the process of α -complementation. Furthermore, methods of mutagenesis, LacZ-alpha variant library construction, selection and characterization (phenotypic and genotypic) were successfully tested.

Key words: protein evolution, genetic code, amino acids, mutagenesis

This thesis is written in Czech.

Obsah

1	SEZNAM POUŽITÝCH ZKRATEK	1
2	TEORETICKÝ ÚVOD	3
2.1	PROTEINOGENNÍ AMINOKYSELINY A JEJICH EVOLUCE	3
2.1.1	<i>Současná podoba genetického kódu.....</i>	4
2.1.2	<i>Evoluce genetického kódu</i>	5
2.1.2.1	Analýza prebiotických podmínek	5
2.1.2.1.1	Exogenní zdroje aminokyselin	6
2.1.2.1.2	Endogenní zdroje aminokyselin	8
2.1.2.2	Analýza současného genomu/proteomu	10
2.1.2.3	Závěry studií evoluce genetického kódu.....	10
2.2	EVOLUČNÍ VZTAH AMINOKYSELINY-PROTEIN	12
2.2.1	<i>Studie náhodných sekvencí.....</i>	12
2.2.2	<i>Redukce souboru aminokyselin u stávajících proteinů.....</i>	13
2.3	POUŽITÍ KNIHOVEN PRO STUDIUM EVOLUCE PROTEINŮ	14
2.3.1	<i>Biologické knihovny.....</i>	14
2.3.1.1	Statistická analýza biologických knihoven	16
2.4	VÝBĚR POŽADOVANÉHO FENOTYPU	18
2.4.1	<i>Lac operon</i>	18
2.4.2	<i>β-galaktosidasa (E.C. 3.2.1.23).....</i>	18
2.4.2.1	Funkce β -galaktosidasy	18
2.4.2.2	Struktura β -galaktosidasy	19
2.4.3	<i>α-komplementace</i>	21
2.4.3.1	Plasmid pUC19	22
2.4.4	<i>Modro/bílá selekce (z angl. Blue/White Screening)</i>	22
3	CÍLE PRÁCE	24
4	MATERIÁL A METODY	25
4.1	CHEMIKÁLIE	25
4.2	BAKTERIÁLNÍ KMENY	26
4.3	PRIMERY	26
4.4	PŘÍSTROJE.....	26
4.5	METODY	27
4.5.1	<i>Polymerázová řetězová reakce (PCR)</i>	27
4.5.2	<i>Mutagenese řízená dvojicí komplementárních primerů (angl. QuickChange PCR).....</i>	28
4.5.3	<i>Elektroforéza v agarosovém gelu</i>	29
4.5.4	<i>Purifikace z gelu</i>	30
4.5.5	<i>Restrikční štěpení.....</i>	30
4.5.6	<i>Ligace</i>	31
4.5.7	<i>Transformace</i>	31
4.5.8	<i>Minipreparace plasmidové DNA.....</i>	32
4.5.9	<i>Sekvenace DNA</i>	33
5	VÝSLEDKY	34
5.1	OVĚŘENÍ FUNKČNOSTI ZKRÁCENÉ VERZE LACZ-ALFA KOMPLEMENTAČNÍHO PEPTIDU.....	34
5.2	KONSTRUKCE KNIHOVEN - ZAVEDENÍ MUTACE.....	37
5.3	VYHODNOCENÍ.....	41

5.3.1	<i>Fenotypizace knihoven</i>	41
5.3.2	<i>Genotypizace knihoven.....</i>	42
5.3.2.1	Knihovna obohacená ranými aminokyselinami.....	43
5.3.2.2	Knihovna obohacená evolučně novějšími aminokyselinami.....	44
6	DISKUSE.....	46
7	ZÁVĚR.....	49
8	SEZNAM POUŽITÉ LITERATURY	50

1 Seznam použitých zkratek

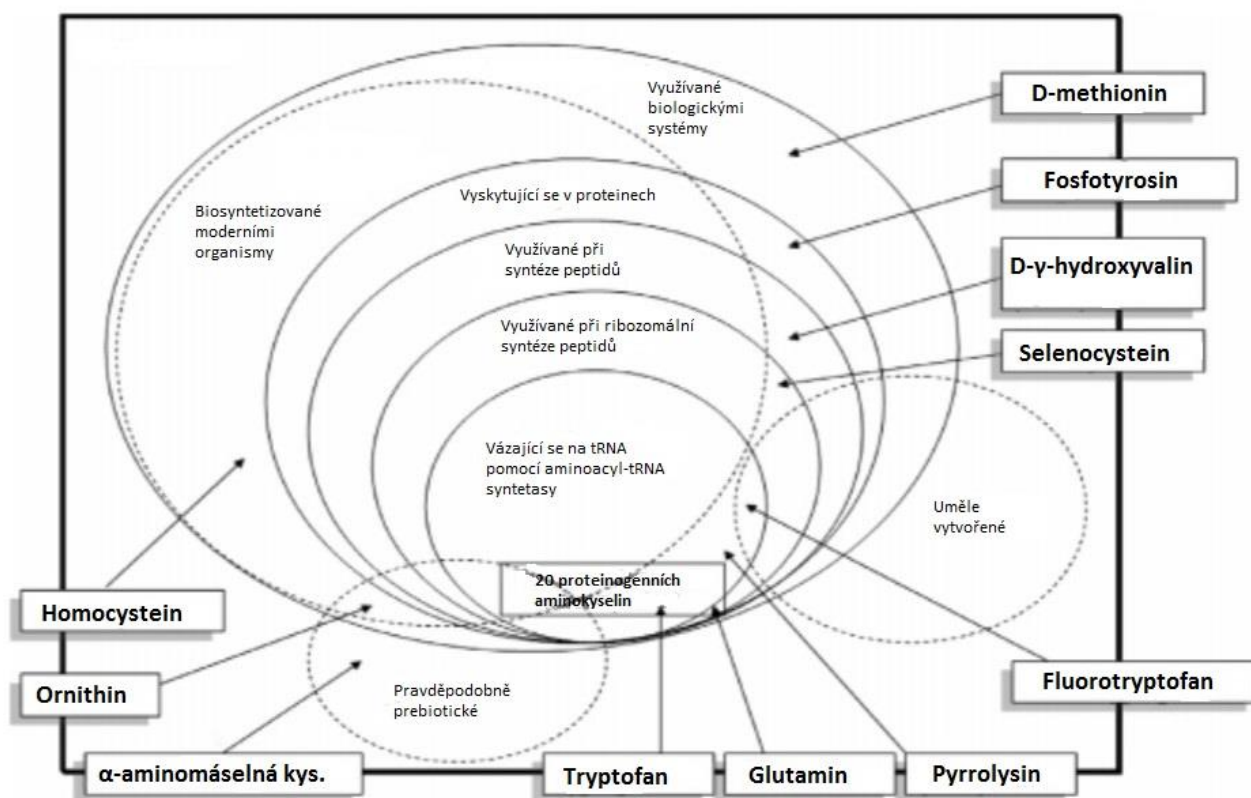
A	adenin
Ala (A)	alanin
Arg (R)	arginin
Asn (N)	asparagin
Asp (D)	kyselina asparagová
bp	pár bazí (z angl. base pair)
C	cytosin
CAP protein	katabolický aktivační protein
Cys (C)	cystein
DMF	dimethylformamid
DNA	deoxyribonukleová kyselina
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	kyselina ethylendiamintetraoctová
FACS	fluorescenční průtoková cytometrie (z angl. fluorescence-activated cell sorting)
G	guanin
Gln (Q)	glutamin
Glu (E)	kyselina glutamová
Gly (G)	glycin
His (H)	histidin
Ile (I)	isoleucin
LB	živné médium podle Luria-Bertaniho
Leu (L)	leucin
LUCA	poslední univerzální společný předek (z angl. Last Universal Common Ancestor)
Lys (K)	lysin
MCS	násobné klonovací místo (z angl. multiple cloning site)
Met (M)	methionin
PCR	polymerázová řetězová reakce (z angl. polymerase chain reaction)
Phe (F)	fenylalanin
Pro (P)	prolin
Pyl	pyrrolysin
RNA	ribonukleová kyselina
RPM	otáčky za minutu (z angl. revolutions per minute)
Sec	selenocystein
Ser (S)	serin
T	thymin
TAE	Tris-acetát-EDTA pufr
TBE	Tris-boritan-EDTA pufr
Thr (T)	threonin
Tris	tris(hydroxymethyl)aminomethan
tRNA	transferová RNA
Trp (W)	tryptofan

Tyr (Y)	tyrosin
U	uracil
UV	ultrafialové záření
Val (V)	valin
w/v	hmotnost/objem (z angl. weight/volume)
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galaktosid

2 Teoretický úvod

2.1 Proteinogenní aminokyseliny a jejich evoluce

Život tak jak ho známe, i přes obrovský repertoár fenotypů, sdílí “centrální dogma molekulární biologie“, jak ho v 50. letech minulého století popsal Francis Crick, spoluobjevitel struktury DNA.^[1] Tedy že organismy genetickou informaci uloženou v DNA přepisují (z angl. transcribe) do RNA a ta je následně přeložena (z angl. translate) na bílkoviny, které vykonávají celou škálu funkcí nezbytných pro metabolismus. Dnes víme, že valná většina dnešních genomů kóduje 20 proteinogenních aminokyselin, které dále podléhají různým posttranslačním modifikacím při tvorbě bílkovin, a také víme, že kromě toho život používá mnoho dalších, geneticky nekódovaných, aminokyselin ve svém metabolismu (Obr. 1).



Obr. 1: Grafické znázornění **výskytu, repertoáru a využití aminokyselin** v biologických systémech. Převzato a upraveno.^[2]

2.1.1 Současná podoba genetického kódu

Genetický kód zajišťuje specifické přiřazování aminokyselin k tripletům nukleotidů a tedy zprostředkovává přenos genetické informace zakódované v RNA. Tři za sebou jdoucí nukleotidy jsou označovány jako kodon a kódováním určité aminokyseliny tak určují i primární strukturu proteinu.^{[3],[4]}

Genetický kód je sestaven z 64 kodonů (Obr. 2 na str. 5), kdy 61 kodonů kóduje 20 proteinogenních aminokyselin a zbylé 3 kodony jsou terminační (označované Stop kodony). Kodon AUG je bifunkční, působí jako iniciační kodon nebo kóduje aminokyselinu methionin. Současný genetický kód je degenerovaný, protože několik redundantních kodonů kóduje jednu aminokyselinu. Tato vlastnost umožňuje výskyt tichých (z angl. silent) mutací, tedy změn kodonů, které se neprojeví ve fenotypu (ve funkci polypeptidového řetězce).^[4]

V současné době převažuje názor, že genetický kód není věcí statickou, ale stále se dynamicky mění v průběhu evoluce. Tuto myšlenku podporuje existence odchylek od univerzálního genetického kódu: mitochondrie obratlovců se liší používáním kodonu AUA pro methionin a UGA pro tryptofan; *Mycoplasma capricolum* používá kodon UGA pro tryptofan; u některých obrvených prvoků se vyskytuje kodon UAR (R = A; G) pro glutamin.^[5]

Jiným podpůrným argumentem je existence dalších dvou aminokyselin, které jsou kódovány v současném genetickém kódu a s velkou pravděpodobností k němu byly evolučně přidány až po evolučním oddělení od posledního univerzálního společného předka LUCA (z angl. Last Universal Common Ancestor).^[2]

Jako první byla objevena nestandardní aminokyselina selenocystein (Sec) v myši glutathionperoxidase a ve formiátdehydrogenase u *E. coli*.^[6] Výskyt této aminokyseliny byl později prokázán u archebakterií, eubakterií i eukaryot. Selenocystein se od cysteinu liší pouze záměnou síry za reaktivnější selen. Tato aminokyselina je kódována tripletem UGA, který v univerzálním genetickém kódu plní úlohu terminačního kodonu. Mezi standardních 20 proteinogenních aminokyselin není řazena hlavně díky několika odchylkám při její translaci.^[2]

O něco později byla objevena další aminokyselina, pyrrolysin (Pyl), přítomná v monomethylamin methyltransferase u archebakterie *Methanosarcina bakeri*.^[6] Výskyt této aminokyseliny byl později prokázán u mnoha dalších methanogenních archebakterií a také u jediné eubakterie *Desulfotobacterium hafniense*, kdy je tato skutečnost někdy považována za důkaz horizontálních přenosů genů. Pyrrolysin je kódován také terminačním kodonem a to konkrétně UAG.^[2]

První pozice (5'-konec)	Druhá pozice v kodonu				Třetí pozice (3'-konec)
	U	C	A	G	
U	UUU-Phe	UCU-Ser	UAU-Tyr	UGU-Cys	U
	UUC-Phe	UCC-Ser	UAC-Tyr	UGC-Cys	C
	UUA-Leu	UCA-Ser	UAA-Ter	UGA-Sec (Ter)	A
	UUG-Leu	UCG-Ser	UAG-Pyl (Ter)	UGG-Trp	G
C	CUU-Leu	CCU-Pro	CAU-His	CGU-Arg	U
	CUC-Leu	CCC-Pro	CAC-His	CGC-Arg	C
	CUA-Leu	CCA-Pro	CAA-Gln	CGA-Arg	A
	CUG-Leu	CCG-Pro	CAG-Gln	CGG-Arg	G
A	AUU-Ile	ACU-Thr	AAU-Asn	AGU-Ser	U
	AUC-Ile	ACC-Thr	AAC-Asn	AGC-Ser	C
	AUA-Ile	ACA-Thr	AAA-Lys	AGA-Arg	A
	*AUG-Met	ACG-Thr	AAG-Lys	AGG-Arg	G
G	GUU-Val	GCU-Ala	GAU-Asp	GGU-Gly	U
	GUC-Val	GCC-Ala	GAC-Asp	GGC-Gly	C
	GUA-Val	GCA-Ala	GAA-Glu	GGA-Gly	A
	GUG-Val	GCG-Ala	GAG-Glu	GGG-Gly	G

Obr. 2: **Tabulka kodonů**, kde * značí iniciační kodon a Ter značí terminační kodon. Převzato a upraveno.^[6]

2.1.2 Evoluce genetického kódu

Ze studia prebiotických podmínek na Zemi a z analýz současných bílkovin a genomů se zdá, že prvotní genetický kód neobsahoval všechny aminokyseliny, které jsou v nynějším genetickém kódu zastoupeny.

2.1.2.1 Analýza prebiotických podmínek

Voda jako médium pro chemické reakce a organické polymery jako nosiče centrální biologické funkce replikace a katalýzy jsou požadavky pro život jak ho známe. Při vzniku Země byla teplota tak vysoká, že přítomnost kapalně vody a zásob organického uhlíku byla v podstatě nemožná. Postupným ochlazováním zemské kůry mohlo začít docházet k postupnému hromadění vody a jednoduchých organických sloučenin, jejichž zdroje

molekul můžeme rozdělit na ty, které byly dodány mimozemskými objekty (exogenní) a na ty vzniklé organickou syntézou poháněnou různými zdroji energie (endogenní).^{[7],[8]}

Tato práce se zabývá aminokyselinami a jimi tvořenými bílkovinami, a proto je tato kapitola ze všech možných prebiotických organických molekul zaměřena pouze na aminokyseliny.

2.1.2.1.1 Exogenní zdroje aminokyselin

S velkou pravděpodobností bylo významné množství aminokyselin dodáno na Zemi mimozemskými objekty jako kometami, meteority (zejména uhlíkatými chondrity) a meziplanetárními prachovými částicemi.^{[8],[9]} Diskutuje se však o možnosti, že nositeli aminokyselin jsou jen meziplanetární prachové částice, které jsou následně začleněny do komet a meteoritů.^[10]

Nejvíce zkoumanými mimozemskými objekty jsou meteority typu uhlíkatých chondritů, které obsahují uhlík jak v elementární formě, tak ve formě sloučenin. První velká analýza aminokyselin byla provedena po dopadu meteoritu Murchison, v němž bylo detekováno 22 aminokyselin, z toho 8 bylo proteinogenních (Obr. 3 na str. 7).^[11] Od té doby bylo analyzováno mnoho dalších meteoritů a s postupným zdokonalováním detekčních metod bylo množství identifikovaných aminokyselin rozšířeno na více než 80. Identifikované aminokyseliny obsahují 2 až 9 atomů uhlíku, kdy s vyšším počtem atomů uhlíků výrazně klesá četnost výskytu. Jako důkaz mimozemského původu těchto aminokyselin svědčí jejich enantiomerní poměr a odlišnost od zastoupení aminokyselin v zemských sedimentech. Z celkového množství identifikovaných aminokyselin bylo zhruba jen 10 proteinogenních (Obr. 3 na str. 7 a Obr. 4 na str. 9).^[12]

C2

Glycin

C3

Alanin

 β -alanin

Serin

Sarkosin

C4

Threonin

allo-threonin α -aminomáselná kyselina β -aminomáselná kyselina α -aminoisomáselná kyselina β -aminoisomáselná kyselina*N*-ethylglycin*N,N*-dimethylglycin*N*-methylalanin*N*-methyl- β -alanin

Asparagová kyselina

2,3-diaminomáselná kyselina

2,4-diaminomáselná kyselina

3,3'-diaminoisomáselná kyselina

C5

Valin

Isovalin

Norvalin

3-aminovalerová kyselina

3-amino-2-methylomáselná kyselina

allo-3-amino-2-methylomáselná kyselina

3-amino-2,2-dimethylpropionová kyselina

3-amino-2-ethylpropionová kyselina

4-aminovalerová kyselina

4-amino-2-methylomáselná kyselina

4-amino-3-methylomáselná kyselina

5-aminovalerová kyselina

Glutamová kyselina

2-methylasparagová kyselina

3-methylasparagová kyselina

allo-3-methylasparagová kyselina*N*-methylasparagová kyselina

4,4'-diaminoisovalerová kyselina

C6

Leucin

Isoleucin

allo-isoleucin

Norleucin

Pseudoleucin

Cyklroleucin

2-methylnorvalin

Piperidin-2-karboxylová kyselina

2-amino-2-ethylomáselná kyselina

2-amino-2,3-dimethylomáselná kyselina

3-amino-2-ethylomáselná kyselina

3-amino-2,3-dimethylomáselná kyselina

3-methylaminvalerová kyselina

4-aminokapronová kyselina

4-amino-3,3-dimethylomáselná kyselina

4-amino-2-methylvalerová kyselina

4-amino-3-methylvalerová kyselina

4-amino-4-methylvalerová kyselina

6-aminokapronová kyselina

 α -aminoadipová kyselina β -aminoadipová kyselina

2-methylglutamová kyselina

ostatní izomery byly detekovány ale nebyly přesně identifikovány

C7

2-amino-2,3,3-trimethylomáselná kyselina

2-amino-2-ethyl-3-methylomáselná kyselina

2-amino-2-ethylvalerová kyselina

2-amino-3-ethylvalerová kyselina

2-amino-2,3-dimethylvalerová kyselina

2-amino-2,4-dimethylvalerová kyselina

2-amino-3,3-dimethylvalerová kyselina

2-amino-3,4-dimethylvalerová kyselina

2-amino-4,4-dimethylvalerová kyselina

allo-2-amino-2,3-dimethylvalerová kys.*allo*-2-amino-3,4-dimethylvalerová kys.

2-amino-2-methylkapronová kyselina

2-amino-3-methylkapronová kyselina

allo-2-amino-3-methylkapronová kyselina

2-amino-4-methylkapronová kyselina

allo-2-amino-4-methylkapronová kyselina

2-amino-5-methylkapronová kyselina

2-aminoenanthová kyselina

 α -aminopimelová kyselina

1-aminocyklohexanová kyselina

ostatní izomery byly detekovány ale nebyly přesně identifikovány

C8

izomery byly detekovány ale nebyly přesně identifikovány

C9

Fenylalanin

Tyrosin

ostatní izomery byly detekovány ale nebyly přesně identifikovány

Obr. 3: Seznam aminokyselin detekovaných na meteoritech. Převzato a upraveno.^[12]

2.1.2.1.2 Endogenní zdroje aminokyselin

Endogenní aminokyseliny mohly být syntetizovány z několika zdrojů, z nichž nejpravděpodobnější byly (i) plynné směsi, (ii) podmořská hydrotermální zřídla (angl. hydrothermal vents), kdy voda působila jako rozpouštědlo a teplota byla nad 100°C a (iii) kapalně směsi, kdy voda působila jako rozpouštědlo a teplota byla pod 100°C.^[13]

Jeden z nejvýznamnějších experimentů simulující prebiotické podmínky na Zemi je experiment Millera a Ureye. Výsledky tohoto experimentu, kterým se pokusili dát vznik organickým sloučeninám za možných primitivních podmínek na Zemi, uveřejnil Stanley L. Miller v roce 1953.^[14] Vycházeli z teorie Oparina a Haldanea, podle které byly základy života položeny v době, kdy atmosféra Země obsahovala metan, amoniak, vodu a vodík.^[15] Byla sestavena aparatura simulující atmosféru tohoto složení, která zde cirkulovala přes elektrické výboje. Co se aminokyselin týče, byl ve výsledné směsi pomocí papírové chromatografie identifikován glycin, α -alanin, β -alanin, kyselina asparagová a α -amino-*n*-másečná kyselina.^[14] Vysoká koncentrace amoniaku v atmosféře byla později zpochybněna a v dalších pokusech nahrazována dusíkem, kdy amoniak byl přítomen jen ve stopovém množství (větší roli mohl amoniak hrát spíše v prebiotických oceánech). Z výbojů v takové experimentální atmosféře bylo za použití plynové chromatografie v kombinaci s hmotnostní spektrometrií v roce 1972 detekováno 10 proteinogenních aminokyselin (Obr. 4 na str. 9, sloupec A2).^[16] Jedním z nejvíce diskutovaných faktorů Millerova experimentu je použití redukční atmosféry. Samotný Miller později provedl svůj experiment tak, aby napodobil podmínky při erupci vulkánu či gejzíru. Do jisker elektrického výboje byly tryskou vstříkovány proudy páry a plynu. Z výsledné směsi bylo identifikováno pět biogenních aminokyselin.^[17] V roce 2008 byla publikována nová analýza, kdy za využití modernějších detekčních metod byly analyzovány téže vzorky z tohoto původního experimentu. Bylo identifikováno 22 aminokyselin, z nichž 7 bylo proteinogenních.^[18] Další experimenty byly provedeny za přítomnosti sulfanu v atmosféře, které vedly k identifikaci aminokyselin jako threonin, leucin a isoleucin. Byla také zkoumána méně redukující atmosféra obsahující oxid uhelnatý a dusík, ve které vznikaly serin, kyselina glutamová, glycin a alanin (Obr. 4 na str. 9, sloupec A3). Výsledky tohoto experimentu také poukázaly na to, jak důležitou prebiotickou podmínkou je voda, jejíž vlastnosti jako pH, či teplota a koncentrace rozpuštěných solí, pravděpodobně ovlivnily prebiotickou evoluci.^[17]

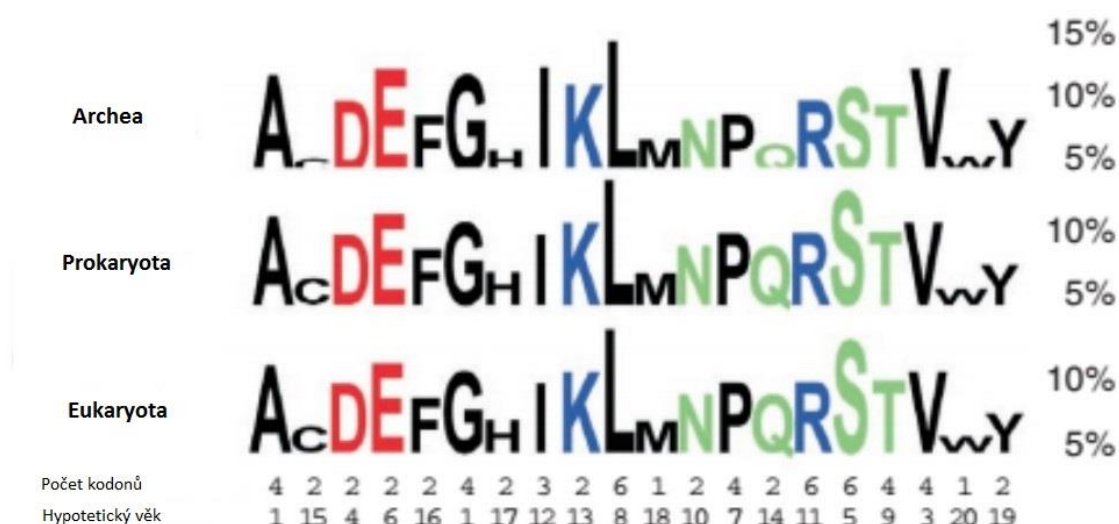
Toto jsou jen nejzákladnější popisy experimentů zkoumajících vznik možných endogenních aminokyselin v prebiotickém období. Pro přehlednost jsou nejvýznamnější endogenní a exogenní zdroje a z nich identifikované aminokyseliny shrnuty na Obr. 4.

	M1	M2	M3	I1	A1	A2	A3	H1	H2	S1	S2	S3	R _{obs}
G - Gly	1.00	1.0	1.000	1.000	1.000	1.000	1.000	18	12	1.000	1.000	40	1.1
A - Ala	0.34	0.4	0.380	0.293	0.540	1.795	0.155	15	8	0.473	0.097	20	2.8
D - Asp	0.19	0.5	0.035	0.022	0.006	0.077	0.059	10	10	0.000	0.581	30	4.3
E - Glu	0.40	0.5	0.110	0.000	0.010	0.018	0.000	6	11	0.000	0.000	20	6.8
V - Val	0.19	0.3	0.100	0.012	0.000	0.044	0.000	1	0	0.006	0.000	2	8.5
S - Ser	0.00	0.0	0.003	0.072	0.000	0.011	0.018	8	11	0.000	0.154	0	8.6
I - Ile	0.13	0.0	0.060	0.000	0.000	0.011	0.000	8	9	0.000	0.002	4	9.1
L - Leu	0.04	0.0	0.035	0.000	0.000	0.026	0.000	3	0	0.001	0.002	7	9.4
P - Pro	0.29	0.1	0.000	0.001	0.000	0.003	0.000	9	0	0.000	0.000	2	10.0
T - Thr	0.00	0.0	0.003	0.000	0.000	0.002	0.000	2	0	0.000	0.002	1	11.7
K - Lys	0.00	0.0	0.000	0.000	0.000	0.000	0.000	7	0	0.000	0.000	14	12.6
F - Phe	0.00	0.0	0.000	0.000	0.000	0.000	0.000	4	0	0.000	0.000	1	13.2
R - Arg	0.00	0.0	0.000	0.000	0.000	0.000	0.000	0	0	0.000	0.000	15	13.3
H - His	0.00	0.0	0.000	0.000	0.000	0.000	0.000	0	0	0.000	0.000	15	13.3
NQCYMW	0.00	0.0	0.000	0.000	0.000	0.000	0.000	0	0	0.000	0.000	0	14.2

Obr. 4: **Relativní koncentrace aminokyselin** pozorované v různých exogenních a endogenních zdrojích, kde M značí meteorit (M1 – Murchison; M2 – Murray; M3 – Yamato), I1 značí meziplanetární částice, A jsou atmosferické syntézy (A1 – CH₄, NH₃, H₂O, H₂; A2 – CH₄, N₂, H₂O, NH₃; A3 – CO, N₂, H₂O), H jsou hydrotermální syntézy a S jsou další chemické syntézy. R_{obs} značí průměrné pořadí vyplývající z těchto výsledků. Ve sloupcích H1, H2 a S3 čísla uvádí počet pokusů, ve kterých byla daná aminokyselina pozorována ve větším než stopovém množství. V ostatních sloupcích čísla značí koncentraci normalizovanou na Gly = 1. Aminokyseliny uvedené na posledním řádku nebyly vůbec v těchto pokusech pozorovány.^[19]

2.1.2.2 Analýza současného genomu/proteomu

Jiný přístup k výzkumu také potvrzuje existenci rané a menší aminokyselinové abecedy. Jsou to současné analýzy metabolických cest, molekul proteinové translace (např. tRNA) a aminokyselinového složení současných genomů (Obr. 5).^{[20],[21]}



Obr. 5: **Zastoupení aminokyselin** u archeobakterií, prokaryot a eukaryot. Výška písmen je přímo úměrná četnosti výskytu odpovídajících aminokyselin. Hypotetický věk značí "stáří" jednotlivých aminokyselin podle Trifonova, kde 1 je nejstarší a 20 nejmladší.^[22] Převzato a upraveno.^[23]

2.1.2.3 Závěry studií evoluce genetického kódu

Z prebiotických podmínek a současných analýz vyplývá, že v prebiotickém světě bylo zřejmě dostupné mnohem větší množství aminokyselin, než je v současné době biologicky aktivních. Velká pozornost je upřena na skutečnost, že přibližně jen polovina ze současných proteinogenních aminokyselin byla přítomna v prebiotickém světě (Obr. 1 na str. 3).

Z mnoha desítek studií byly provedeny dvě velké metaanalýzy. Asi největší analýza byla uveřejněna v roce 2000 E. N. Trifonovem. Ten do své analýzy zahrnul 40 různých kritérií a hypotéz o vzniku a evoluci genetického kódu k vytvoření časového pořadí výskytu aminokyselin v průběhu evoluce. Výsledky analýzy souhlasily s výsledky experimentů S. L. Millera (i z hlediska pořadí) a chronologicky zorganizovaly všech 64 kodonů se zjištěním, že nové kodony přibývaly do genetického kódu v pořadí odpovídajícímu jejich klesající

termostabilitě.^[22] O 4 roky později Trifonov rozšířil počet kritérií na 60 a ještě více upřesnil pořadí chronologického přidávání aminokyselin do genetického kódu (Obr. 6).^[24]

	U	C	A	G		
U	Phe	Ser	Tyr	Cys	U	
	Leu		Stop	Trp	C	
					A	
C	Leu	Pro	His	Arg	G	
			Gln		U	
					C	
					A	
A	Ile	Thr	Asn	Ser	G	
	Met		Lys		Arg	U
						C
						A
G	Val	Ala	Asp	Gly	G	
			Glu		U	
					C	

Průměrné pořadí

< 6
6 - 9
9 - 12
> 12

Obr. 6: Tabulka **univerzálního genetického kódu** s rozdělenými aminokyselinami podle jejich průměrného pořadí přidávání do genetického kódu vypočteného Trifonovem podle 60 kritérií. Horní polovina tabulky odpovídá kodonu YNN (Y = U, C; N = U, C, A, G) a je obohacena evolučně novějšími aminokyselinami, zatímco dolní polovina tabulky odpovídá kodonu RNN (R = A, G) a je obohacena ranými aminokyselinami. Převzato a upraveno.^[25]

V roce 2009 byla Paulem G. Higgsem a Ralphem E. Pudritzem uveřejněna metaanalýza podobná Trifonovově, ale některá kritéria či hypotézy byly ze souboru vyřazeny a byly zde zahrnuty pouze studie kvantifikující výskyt aminokyselin.^[19]

Nicméně se obě tyto metaanalýzy shodují ve svém závěru, že prvotní genetický kód obsahoval asi jen polovinu nynějších proteinogenních aminokyselin, které jsou v této práci považovány za rané. Je to **glycin (Gly)**, **alanin (Ala)**, **kyselina asparagová (Asp)**, **kyselina glutamová (Glu)**, **valin (Val)**, **isoleucin (Ile)**, **leucin (Leu)**, **prolin (Pro)**, **threonin (Thr)** a **serin (Ser)**.

V literatuře se často uvádí, že bazické aminokyseliny nebyly podle žádné z dosavadních studií prebioticky dostupné a kódované původním genetickým kódem. Bylo prozkoumáno několik moderních genových sekvencí kódujících známé i předpokládané proteiny, které neobsahovaly arginin a lysin a některé z těchto proteinů mají známou a důležitou funkci (např. měď vázající protein (z angl. copper ion binding protein)). Z toho

vyplývá, že nedostatek bazických aminokyselin nebrání poskládání proteinu do stabilní trojrozměrné struktury (z angl. folding) a enzymové aktivitě proteinu. Otázkou zůstává, zda a jak mohly tyto předpokládané prvotní proteiny bez lysinu a argininu interagovat s nukleovými kyselinami.^[26] Nelze rovněž vyloučit, že původní kód obsahoval i jiné, prebioticky spíše dostupné aminokyseliny, které byly postupně nahrazeny současnými.

Z porovnání faktorů, které se podílely na selekci 20ti proteinogenních aminokyselin, byla vyslovena zajímavá domněnka. Pokud by měl život nastat na jiné planetě, bylo by 75% proteinogenních aminokyselin stejných a bylo by velmi překvapující, kdyby jich bylo o 5 méně či více stejných.^[27] Dnes se někteří domnívají, že by nebylo překvapující, kdyby extraterestriální genetický kód obsahoval 20 stejných proteinogenních aminokyselin.^[9]

2.2 Evoluční vztah aminokyseliny-protein

Studie z oboru evoluční biologie a proteinového inženýrství poukazují na to, že prvotní proteiny mohly být tvořeny menším počtem proteinogenních aminokyselin a přesto poskytovaly dostatečný strukturní a funkční repertoár. Tyto studie se zabývají jednak proteiny, u kterých byl *in vitro* evolucí redukován soubor aminokyselin a za druhé náhodnými sekvencemi tvořenými různými podmnožinami aminokyselin.

2.2.1 Studie náhodných sekvencí

Skupina Hiroshiho Yanagawy pracovala s knihovnami proteinů složených z redukovaného souboru aminokyselin a porovnávala je s proteiny složenými z 20ti proteinogenních aminokyselin. Tyto knihovny byly kódovány pomocí kodonů GNN kódující alanin, glycin, kyselinu asparagovou, kyselinu glutamovou a valin (podle většiny analýz prebiotických podmínek to jsou nejhojněji se vyskytující prebiotické aminokyseliny); RNN (R = A, G) kódující v porovnání s GNN navíc isoleucin, methionin, threonin, asparagin, lysin, serin a arginin (významné je zde zejména obohacení o evolučně novější aminokyseliny a s tím i výskyt kladně nabitých aminokyselin) a NNN kódující všech 20 proteinogenních aminokyselin. Takto získané proteiny se lišily svou rozpustností. „GNN proteiny“ vykazovaly nejvyšší rozpustnost, „RNN proteiny“ vykazovaly střední rozpustnost a jako nejméně rozpustné se jevily „NNN proteiny“.^{[25],[28],[29]}

Z jejich studií mimo jiné vyplývá, že knihovny náhodných sekvencí složené z raných aminokyselin by mohly být pro jejich lepší rozpustnost vhodnějším zdrojem pro hledání bílkovin a peptidů s novými funkcemi a strukturami. I tato skupina se dále zabývala redukcí souboru aminokyselin, z nichž mohou být sestaveny funkční bílkoviny (kap. 2.2.2) a na modelovém SH3 proteinu potvrdili, že je snadnější redukovat soubor aminokyselin u stávajících proteinů pomocí raných aminokyselin než pomocí evolučně nových aminokyselin.^[30]

2.2.2 Redukce souboru aminokyselin u stávajících proteinů

Jednou z velkých otázek v oboru proteinového inženýrství je, kolik různých aminokyselin je zapotřebí pro tvorbu sbalených bílkovin a z jakých aminokyselin se takové redukované sekvence mohou skládat.^[31] Úspěšně byly vyprodukovány proteiny skládající se z 5 - 10 z celkových 20ti proteinogenních aminokyselin a u mnohých z nich byla v plném rozsahu zachována jejich vazebná či katalytická funkce.^{[30],[32]–[40]} Strukturní podoba těchto bílkovin byla různá – od malých proteinů tvořených β -skládanými listy (např. SH3 domény), přes α -helikální bílkoviny (angl. four helix bundle) až po $\alpha\beta\beta$ strukturní motivy.^{[32],[36],[38]}

Jen některé z výše popsaných studií zároveň zahrnuly evoluci genetického kódu a zabývaly se bílkoviny tvořenými ranými aminokyselinami. Všechny tyto studie se však omezily na obohacení bílkovin ranými aminokyselinami nanejvýš na 80%.^{[30],[33]}

Pomocí *in silico* studie došli Babajide et al. k závěru, že struktury podobající se nativním formám sbalených proteinů se na několika testovaných bílkovinách daří zachovat spíše pomocí omezeného souboru aminokyselin tvořených ranými aminokyselinami než pomocí souboru tvořených evolučně novějšími aminokyselinami.^[41]

Práce popsaná níže si klade za cíl připravit metodiku, která by umožnila ověřit na vybraném proteinu možnosti záměny evolučně novějších aminokyselin za rané za současného udržení jeho strukturně-funkční integrity.

2.3 Použití knihoven pro studium evoluce proteinů

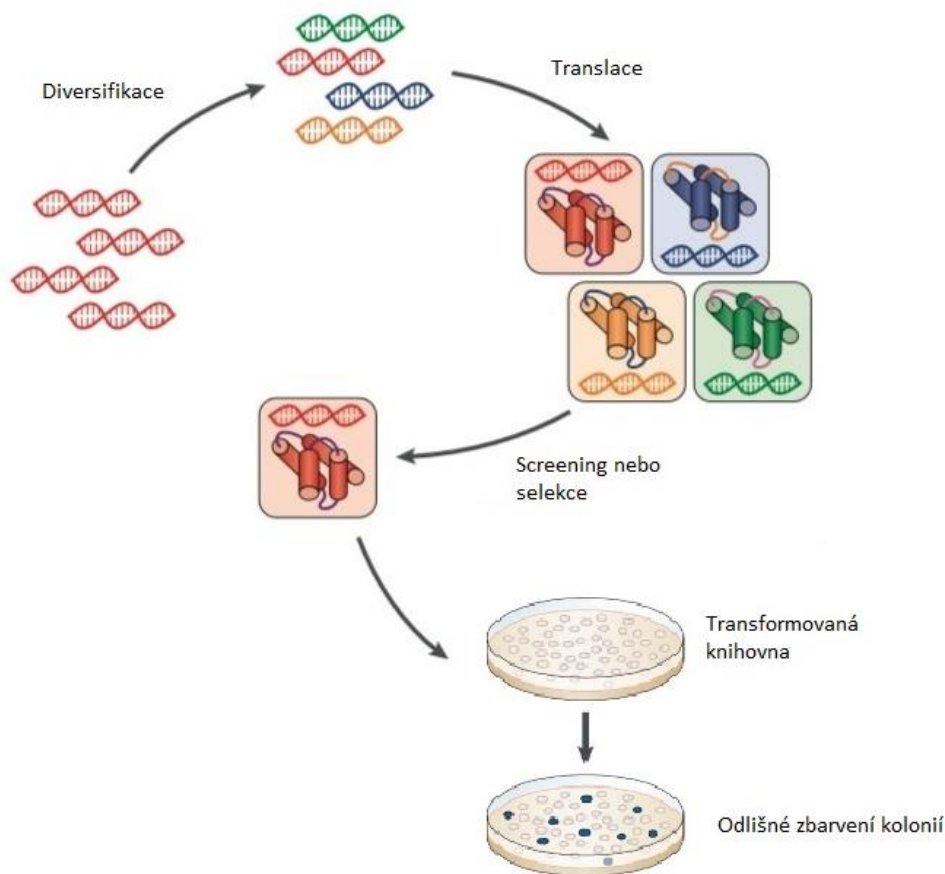
Účinným nástrojem pro *in vitro* evoluci proteinů jsou biologické knihovny, které představují soubor populací fragmentů DNA. Po jejich expresi ve vhodném systému jsou podrobeny individuální analýze (angl. screening) nebo selekci za účelem získání souboru variant požadovaných vlastností.^[42]

2.3.1 Biologické knihovny

Biologické knihovny obvykle pracují s DNA nebo RNA kódujícími genetické varianty studovaného proteinu. Prvním krokem je diversifikace genu. Provádí se buď náhodnou mutagenezí, kdy jsou do náhodných míst zaváděny náhodné mutace, nebo místně specifickou mutagenezí, kdy jsou mutovány či randomizovány vybrané aminokyselinové pozice. Někdy je používána diversifikace rekombinací, kdy se kombinují varianty celých genových úseků. Získaný soubor genetických variant se exprimuje ve vhodném systému (bakterie, eukaryotní buňky, *in vitro* translace) a je sledován jeho fenotyp (Obr. 7 na str. 15).^[42] Protože *in vitro* translační systém není závislý na transformaci žádného hostitele, je množství jednotlivých zkoumaných variant sekvencí (10^{14}) výrazně vyšší než pro *in vivo* systémy (obvykle $10^6 - 10^8$ variant). Proto jsou *in vitro* translační systémy obzvláště vhodné pro zkoumání velkého sekvenčního prostoru.^[43]

Jedním typem experimentálního uspořádání *in vivo* je, že každá jednotlivá expresní jednotka (např. bakteriální kolonie), obsahující právě jednu variantu zkoumaného proteinu, je sledována a vyhodnocována individuálně. Podle toho, jak jednotlivé varianty proteinu ovlivní fenotyp (např. barvu kolonie), lze najednou prozkoumat celý soubor mutovaných proteinů a vybrat varianty požadovaných vlastností. Nevýhodou je, že touto cestou lze prozkoumat relativně malý počet možností, výhodou je možnost aplikace některé z mnoha metod fenotypování. Částečně zautomatizovaným sledováním fenotypu je typ průtokové cytometrie FACS (z angl. fluorescence-activated cell sorting), kdy při dopadu laserového paprsku dochází k třídění buněk podle jejich fluorescence. Příkladem je využití zeleného fluorescenčního proteinu, jehož přítomnost či absence ve variantách proteinu je zjišťována právě fluorescencí. Metoda použitá v této bakalářské práci je založena na modro/bílé selekci

(z angl. blue/white screening), kdy modré kolonie indikují přítomnost aktivního enzymu β -galaktosidasy, zatímco bílé kolonie značí přítomnost neaktivní formy tohoto enzymu.^[42]



Obr. 7: Příklad experimentálního uspořádání při konstrukci a analýze **biologických knihoven**. Převzato a upraveno.^[42]

Při selekci není potřeba analyzovat každou jednotlivou variantu zvlášť, místo toho je využita fyzická separace celé populace variant, které mají požadovanou vlastnost. Mezi základní selekční metody patří selekce.^[42]

(i) na základě vazebné afinity

Varianty jednoho interakčního partnera se exprimují na povrchu bakteriofága pomocí fúze s jeho vlastním povrchovým proteinem a na základě vazebné afinity k imobilizovanému cílovému interakčnímu partnerovi je lze selektovat. Ze zachyceného fága lze izolovat odpovídající DNA.^[44]

(ii) na základě přežití organismu

Nejjednodušším příkladem selekce na základě přežití organismu je rezistence vůči antibiotikům. Dalším příkladem může být auxotrofní komplementace, kdy funkční varianty získané z knihovny jsou schopné kompenzovat určitý metabolický defekt hostitelské buňky.^[42]

2.3.1.1 Statistická analýza biologických knihoven

K analýze úplnosti a rozmanitosti v knihovnách, které vznikly mutagenézí oligonukleotidů, se používají statistické metody. Umožňují předpovědět potřebnou velikost knihovny tak, aby pravděpodobnost, že všechny možnosti budou zastoupeny, byla dostatečně vysoká. Logicky lze předpokládat, že čím větší počet kodonů bude mutován, tím menší bude pravděpodobnost získání všech možných randomizovaných variant. Velkým problémem je, že většinu teoretických úvah není možné experimentálně prověřit. V pozorovaném Vesmíru je přibližně 10^{80} atomů, což je zanedbatelné číslo vůči počtu možných sekvencí DNA v proteinu. I malý protein, složený ze sta aminokyselin, může být kódován přibližně 10^{181} možnými sekvencemi DNA.

Počet studovaných klonů (např. počet získaných klonů) označme L a počet všech možných variant sekvence V .

K zjištění počtu odlišných variant v knihovně vycházíme z Poissonova rozdělení pravděpodobnosti, že $L \gg V$.

Poissonova distribuční rovnice:

$$P_{(x)} = \frac{e^{-\lambda} \lambda^x}{x!} \quad (1)$$

kde $P_{(x)}$ značí pravděpodobnost, že se určitá varianta v_i objeví právě x -krát v dané knihovně a λ značí parametr Poissonova rozdělení pravděpodobnosti, kdy $\lambda = L/V$, tedy průměrný počet výskytů jakékoliv jedné varianty v_i v knihovně.

Pro pravděpodobnost, že se daná varianta objeví nejméně jednou, platí

$$1 - P(0) = 1 - e^{-\lambda} = 1 - e^{-\frac{L}{V}} \quad (2)$$

a tedy počet všech možných variant C můžeme vyjádřit jako

$$C = V \left(1 - e^{-\frac{L}{V}} \right) \quad (3)$$

Frakční úplnost knihovny je dána

$$F = \frac{C}{V} = 1 - e^{-\frac{L}{V}} \quad (4)$$

a udává procentuální zastoupení variant ze všech možných.

Výpočet potřebné velikosti knihovny, aby byla s danou pravděpodobností kompletní, je o něco složitější. Potřebujeme tedy vypočítat počet klonů L .

Pravděpodobnost, že je zastoupena každá varianta, je vyjádřena jako

$$P_c = (1 - e^{-\frac{L}{V}})^V \quad (5)$$

a po matematických úpravách této rovnice můžeme vyjádřit L jako

$$L = -V \ln\left(-\frac{\ln P_c}{V}\right) \quad (6)$$

V experimentu, popsaném v této bakalářské práci, jsme měli 32 variant kodonů v každé knihovně, tedy $V = 32$ (všech 32 možných kodonů a Stop kodonů).

K sestrojení knihovny, obsahující 95% ($F = 0,95$) všech možných variant, můžeme odhadnout potřebný počet analyzovaných klonů L podle rovnice (4).

$$0,95 = 1 - e^{-\frac{L}{V}} \rightarrow L = -32 \ln 0,05 = 96$$

Potřebný počet klonů L pro 95% pravděpodobnost detekce všech variant vypočítáme podle rovnice (6).

$$L = -V \ln\left(-\frac{\ln P_c}{V}\right) = -32 \ln\left(-\frac{\ln 0,95}{32}\right) = 206$$

Podle těchto orientačních výsledků je pro knihovnu, která pokryje 95% variant, třeba analyzovat 96 klonů a pro 95% pravděpodobnost zisku kompletní knihovny je třeba získat 206 kolonií.

S každou procentuální změnou pravděpodobnosti se čísla výrazně liší. Pro 90% úplnost knihovny je třeba analyzovat 74 klonů a pro 90% pravděpodobnost zisku kompletní knihovny je třeba získat 183 kolonií. Naopak pro 99% vychází 147 klonů a 258 kolonií.^[45] S použitím vhodné fenotypizace lze z knihovny vybrat pouze varianty požadovaného fenotypu, čímž se může výrazně snížit počet selektovaných nebo analyzovaných klonů.

2.4 Výběr požadovaného fenotypu

Metoda fenotypizace použitá v této bakalářské práci je založena na modro/bílé selekci (z angl. blue/white screening), kdy modré kolonie indikují přítomnost aktivního enzymu β -galaktosidasy, zatímco bílé kolonie značí přítomnost neaktivní formy tohoto enzymu.

2.4.1 Lac operon

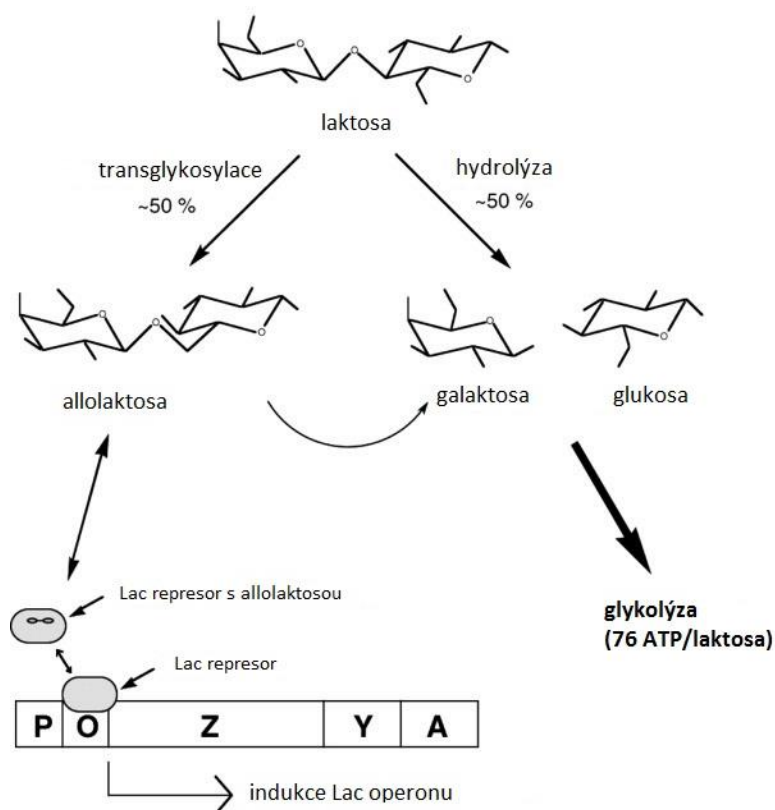
Lac operon *E. coli* je transkripční jednotka, která reguluje a produkuje enzymy potřebné k metabolismu laktosy. Obsahuje tři strukturní geny a to lacZ kódující β -galaktosidasu, lacY kódující β -galaktosidpermeasu a lacA kódující galaktosidacetyltransferasu.

Lac operon je regulován pozitivně i negativně. Pozitivní regulace probíhá prostřednictvím katabolitů CAP proteinu (katabolický aktivační protein), které v nepřítomnosti glukosy působí jako aktivátory transkripce. Negativní regulace probíhá pomocí derivátu laktosy, allolaktosy. Vazba represoru LacI způsobuje zastavení transkripce a právě navázání allolaktosy na represor způsobí jeho inaktivaci. Pro úplnou expresi Lac operonu je tedy nutné, aby jako zdroj uhlíku nebyla přítomna glukosa, ale aby byla přítomna laktosa.^{[46],[47]}

2.4.2 β -galaktosidasa (E.C. 3.2.1.23)

2.4.2.1 Funkce β -galaktosidasy

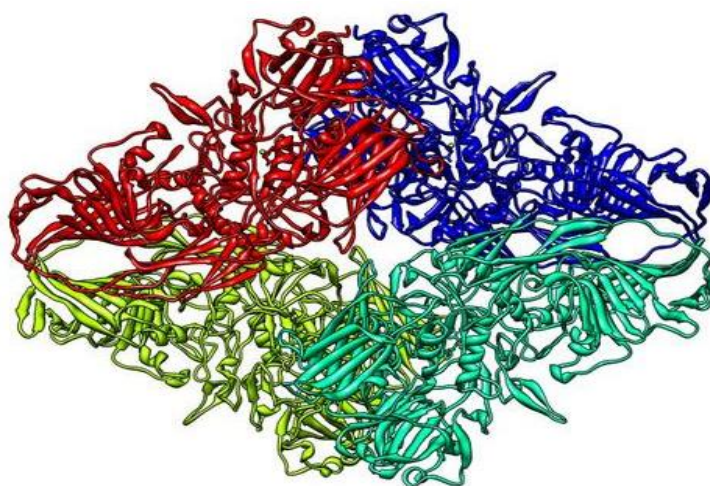
Tento enzym má trojí enzymatickou aktivitu. Štěpí disacharid laktosu na monosacharidy glukosu a galaktosu, katalyzuje transgalaktosylaci laktosy na allolaktosu a štěpí allolaktosu na monosacharidy (Obr. 8 na str. 19).^[47]



Obr. 8: Trojí enzymatická funkce **β -galaktosidasy** a negativní regulace **Lac operonu**. Převzato a upraveno.^[47]

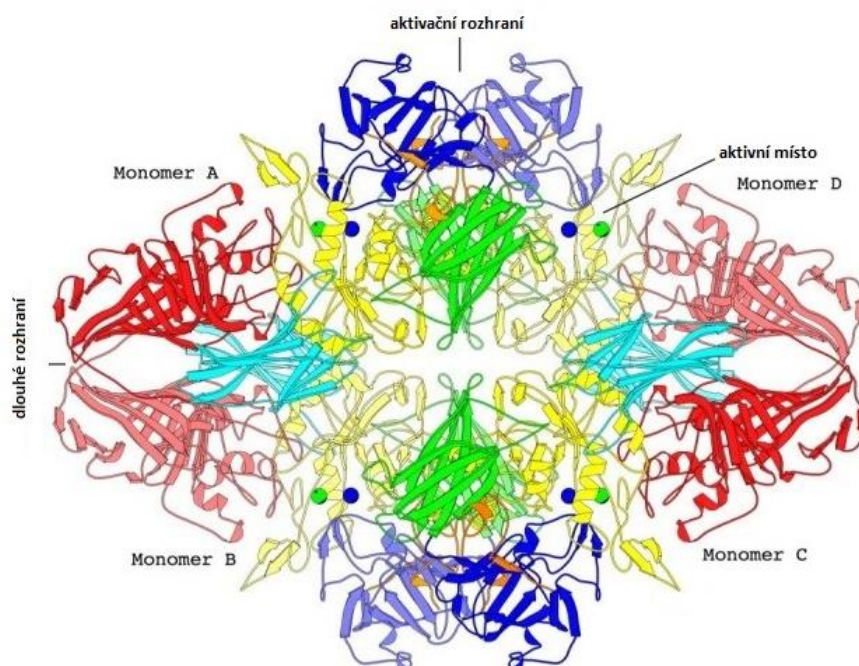
2.4.2.2 Struktura **β -galaktosidasy**

β -galaktosidasa je složena ze čtyř identických monomerů (podjednotek) vyznačených na Obr. 9.



Obr. 9: Struktura **β -galaktosidasy** s barevně rozlišenými monomery.^[48]

Tyto monomery jsou seskupeny okolo tří os symetrie, které tvoří tři odlišná rozhraní (z angl. interface) mezi různými páry monomerů. Dlouhé rozhraní je tvořeno horizontální osou, která odděluje páry monomerů A/B a C/D. Aktivační rozhraní je tvořeno vertikální osou, která odděluje monomery A/D a B/C. Třetí, nejmenší rozhraní tvořené kolmou osou, odděluje monomery A/C a B/D (Obr. 10).



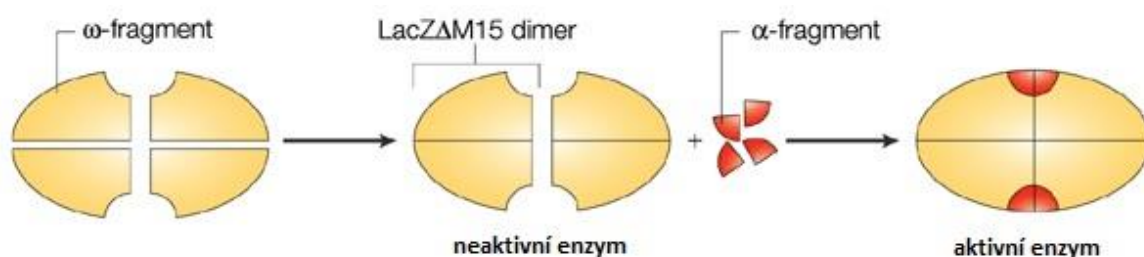
Obr. 10: **Struktura β -galaktosidasy** s barevně rozlišenými doménami (doména 1 – tmavě modrá/fialová; doména 2 - zelená; doména 3 - žlutá; doména 4 - tyrkysová; doména 5 – červená/růžová), kdy jednotlivé odstíny dané barvy značí stejný typ domény dvou různých monomerů. Jako kuličky jsou zde znázorněny sodné (zelená) a hořečnaté (modrá) ionty přítomny na všech čtyřech aktivních místech. Převzato a upraveno.^[49]

Každý monomer je složen z 1023 aminokyselin, které tvoří pět strukturních domén (Obr. 10). Nejdůležitější je třetí doména, označovaná TIM barel (z angl. triose phosphate isomerase barrel), která také bývá označována jako centrální doména díky přítomnosti aktivního místa na C-terminálním konci domény. Aktivní místo je tedy z velké části tvořeno hlubokou prohlubní třetí domény, ale zasahují sem i smyčky druhé a páté domény daného monomeru a také sem přes aktivační rozhraní zasahuje smyčka druhé domény jiného monomeru. Celkem tedy enzym obsahuje čtyři samostatná aktivní místa, avšak z jejich složení je zřejmé, že jednotlivé monomery nemohou být samy o sobě aktivní, protože je potřeba dvou monomerů k úplnosti aktivního místa. Dále musí být k maximální aktivitě β -galaktosidasy přítomny sodné a hořečnaté ionty.^{[47],[49]}

2.4.3 α -komplementace

Delece určitých sekvencí na amino-terminálním konci enzymu (nejčastěji zbytky 23 - 31 nebo 11 - 41) způsobuje disociaci aktivního tetrameru na dva neaktivní dimery nazývané LacZ Δ 15. Dochází tedy k narušení aktivační rozhraní, zatímco dlouhé rozhraní je zachováno a umožňuje existenci dimerů. Tyto dimery jsou nestálé a v případě nedostatku sodných a hořečnatých iontů disociují na monomery.

Fragment enzymu postrádající LacZ-alfa peptid (zbytky 11 - 41) nazýváme LacZ-omega peptid. Využitím různě dlouhého peptidu obsahujícího tyto chybějící zbytky (klasicky 3 - 41 nebo 3 - 92) je možná rekonstrukce aktivního enzymu v procesu nazvaném α -komplementace zobrazeném na Obr. 11.^{[47],[49]}



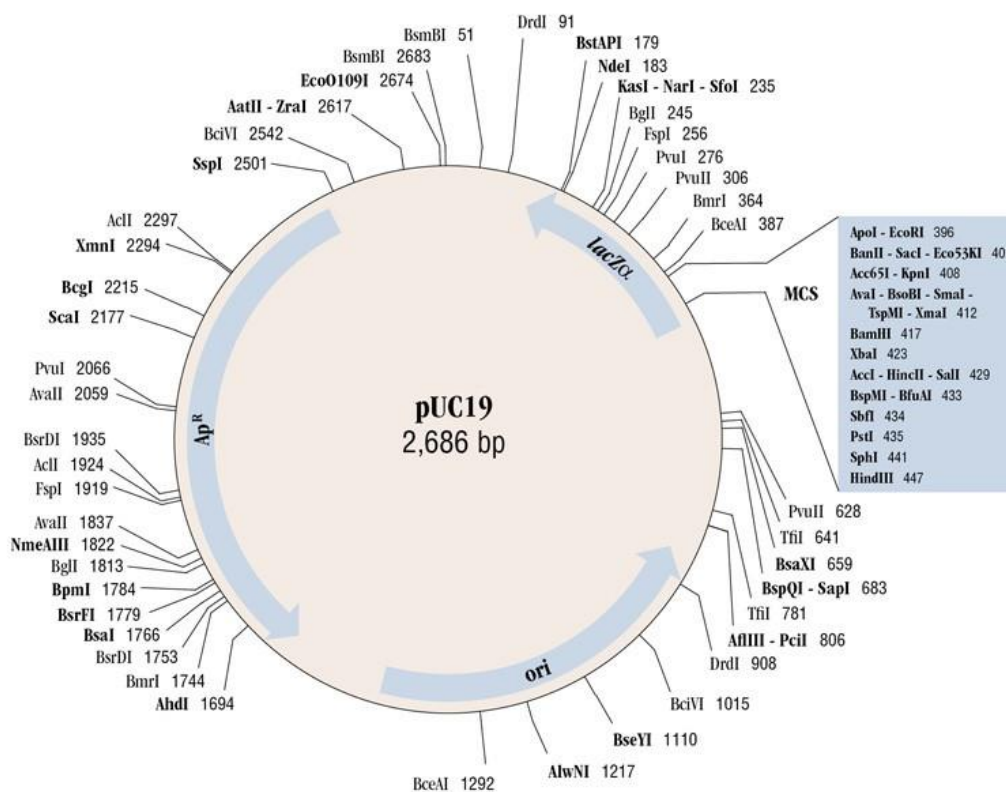
Obr. 11: Obecně znázorněný princip α -komplementace. Převzato a upraveno.^[50]

Z krystalové struktury enzymu a z chromatografického zkoumání bylo odvozeno, že k úspěšné α -komplementaci je nutný poměr ω : α peptidu 1:1. Novější výzkumy, založené na fluorescenční flukтуаční analýze, došly k rozdílnému názoru a to že funkční poměr ω : α peptidu je 4:1.^[51]

V praxi je α -komplementace často využívána u kmenů *E. coli*, které produkují pouze LacZ-omega peptid. Do takového kmene je vnesen plasmid exprimující LacZ-alfa peptid a dojde ke spojení obou podjednotek a k tvorbě aktivní β -galaktosidasy. Aktivitu enzymu lze snadno detekovat pomocí modro/bílé selekce.^{[47],[49]}

2.4.3.1 Plasmid pUC19

Plasmid pUC19 je běžně využívaná dvouřetězcová cirkulární molekula o délce 2686 párů bazí (Obr. 12). Má vysoký počet klonů díky delecí genu kódujícího protein, který kontroluje replikaci. Tento plasmid uděluje resistenci na ampicilin a nese LacZ-alfa peptid, do kterého byl vložen polylinker. Tento polylinker, o délce 54 párů bazí, je násobným klonovacím místem (MCS z angl. multiple cloning site) pro 13 různých restričních endonukleas (např. EcoRI, HindIII, PstI, SacI). Vložení cizorodé DNA do MCS může přerušit LacZa gen a tedy zapříčinit neschopnost katabolismu laktosy.^{[52],[53]}

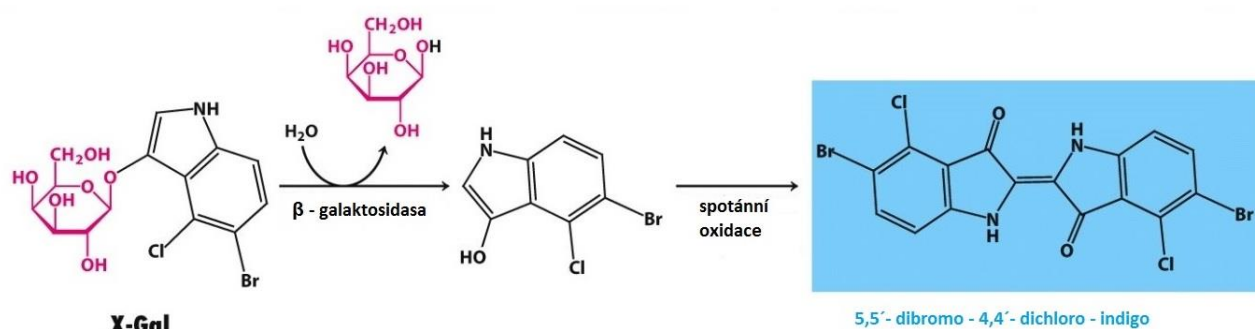


Obr. 12: Genová mapa plasmidu pUC19. Převzato a upraveno.^[54]

2.4.4 Modro/bílá selekce (z angl. Blue/White Screening)

Jak již bylo zmíněno, je tato metoda využívána ke sledování aktivní a neaktivní formy β -galaktosidasy. V našem případě jsme sledovali funkčnost LacZ-alfa peptidu o různé sekvenci aminokyselin po jeho vnesení a expresi v buňkách *E. coli* kódujících LacZ-omega

peptid. Barevná rozdílnost kolonií je způsobena jejich růstem v přítomnosti chromogenního substrátu X-gal (5-bromo-4-chloro-3-indolyl- β -D-galaktosid). Ten je bezbarvou látkou, která je, jako analog laktosy, štěpena β -galaktosidasou na modrý produkt (Obr. 13). Modré kolonie tedy nesou funkční sekvenci LacZ-alfa peptidu a produkují aktivní β -galaktosidasu, zatímco u bílých kolonií byla přítomna nefunkční sekvence.^{[47],[49]}



Obr. 13: Štěpení **chromogenního substrátu X-gal** β -galaktosidasou a následná spontánní reakce vedoucí k modře zbarvenému produktu. Převzato a upraveno.^[55]

3 Cíle práce

- Ověření funkčnosti zkrácené formy LacZ-alfa peptidu β -galaktosidasy pro α -komplementaci.
- Příprava dvou knihoven mutantů LacZ-alfa peptidu metodami místně specifické mutagenese.
- Určení genotypu ve vztahu k fenotypu u vybraných variant.
- Statistická analýza získaných dat.
- Zhodnocení vhodnosti použití LacZ-alfa peptidu a metod mutagenese pro další mikroevoluční studie s použitím α -komplementace.

4 Materiál a metody

4.1 Chemikálie

- Biotium (Hayward, USA)
Fluorescenční barvivo: GelRed
- Invitrogen (Waltham, USA)
Plasmid: pUC 19
- Qiagen (Hilden, Německo)
Kit pro purifikaci z gelu: QIAquick Gel Extraction Kit
Kit pro minipreparaci: QIAprep Spin Miniprep Kit
Marker molekulární hmotnosti DNA: GelPilot 100 bp Plus Ladder
- New England BioLabs (Ipswich, USA)
Pufr pro restrikční štěpení: NEBuffer 2; NEBuffer 2.1.; NEBuffer 3; CutSmart Buffer
Restrikční endonukleasy: EcoRI; NdeI; HindIII; PvuII
Pufr pro ligaci: 10x Buffer for T4 DNA Ligase with 10mM ATP
Ligasa: T4 DNA Ligase
- Penta (Praha, Česká republika)
Isopropylalkohol
- Serva (Heidelberg, Německo)
Agarosa pro DNA elektroforézu
- Top-Bio (Praha, Česká Republika)
Master Mix pro PCR: PPP Master Mix (2x conc. + Mg^{2+})
- ÚOCHB AV ČR, v.v.i. (Praha, Česká republika)
Nanášecí pufr (40% (w/v) sacharosa; 0,1% (w/v) bromfenolová modř)
Elektroforetické pufr: TAE pufr (40mM Tris (pH 7,6); 20mM kyselina octová; 1 mM EDTA); TBE pufr (89mM Tris (pH 7,6); 89mM kyselina boritá; 2mM EDTA)

4.2 Bakteriální kmeny

- Top10 *E. coli* (firma Life Technologies, Praha, Česká Republika)
- Mach1 *E. coli* (firma Life Technologies, Praha, Česká Republika)

4.3 Primery

Všechny primery byly objednány u firmy GENERI BIOTECH s.r.o. (Hradec Králové, Česká Republika)

- **WRNNDown**; R = (A; G)
5'-CACCATATGTTACTCTTCGGAGTTCCGNNRGCTGGCGAAAGGGG-3'
- **WRNNUp**; R = (A; G)
5'-CCCCTTTTCGCCAGCRNNCGGAACTCCGAAGAGTAACATATGGTG-3'
- **WYNNDown**; Y = (T; C)
5'-CACCATATGTTACTCTTCGGAGTTCCGNNYGCTGGCGAAAGGGG-3'
- **WYNNUp**; Y = (T; C)
5'-CCCCTTTTCGCCAGCYNNCGGAACTCCGAAGAGTAACATATGGTG-3'
- **M13 rev. primer**
5'-CGTATGTTGTGTGGAATTGTGAG-3'

4.4 Přístroje

- Centrifugy: Allegra X-15R, Beckman Coulter, Inc. (USA)
Biofuge pico, Heraeus Instrument (Německo)
- UV transiluminátor QUANTUM ST4-1100, Vilber Lourmat (Německo)
- Inkubátor, Memmert (USA)
- Systém pro horizontální elektroforézu Owl, Thermo Scientific (USA)
- Laboratorní váhy HL-400, A&D Engineering, Inc. (USA)
- Spektrofotometr NanoDrop 1000, Thermo Scientific (USA)
- Třepačka Innova 44, New Brunswick Scientific (Německo)
- Zdroj napětí Consort EV243, Sigma-Aldrich (USA)

- Termální cyklátor T Gradient, Biometra (Německo)
- Blokový termostat MB-102, Bioer (Čína)

4.5 Metody

4.5.1 Polymerázová řetězová reakce (PCR)

PCR reakce, využívající teplotní cykly, byla použita k amplifikaci vymezeného úseku DNA plasmidu pUC19 – STOP (kap. 5.1).

Postup PCR:

Do PCR-mikrozkumavky byla pipetována reakční směs o celkovém objemu 39,9 μl podle Tab. 1. Ke kontrolnímu vzorku nebyl přidán templát. Reakce probíhala podle Tab. 2.

Složka	Objem V [μl]
PPP MasterMix	20
Sterilní voda	19
Přímý primer (z angl. forward): M13 rev. primer (100 μM)	0,4
Zpětný primer (z angl. reverse): WRNNDown/WYNNDDown (100 μM)	0,4
Templát pUC19-STOP (10 ng/ μl)	0,1

Tab. 1: Složení reakční směsi pro PCR reakci

Fáze	Teplota [$^{\circ}\text{C}$]	Čas [s]	Počet cyklů
1	95	120	1
2	98	30	32
3	60	30	
4	72	40	

Tab. 2: Průběh fází PCR reakce

4.5.2 Mutageneze řízená dvojicí komplementárních primerů (angl. QuickChange PCR)

Jedná se o modifikaci metody PCR, kdy došlo k vytvoření mutací *in vitro* přímo na plasmidu pUC19 - STOP. Byly použity vzájemně komplementární primery (kap. 4.3). Během QuickChange PCR reakce se syntetizuje celé vlákno plasmidu a výsledný produkt by měl odpovídat jeho linearizované podobě. Mateřská plasmidová DNA je, na rozdíl od DNA syntetizované během PCR reakce, methylována a může být proto odstraněna restriční endonukleasou DpnI. V reakční směsi by poté měly zůstat pouze nemethylované produkty QuickChange PCR reakce.

Postup:

Do sterilní PCR-mikrozkumavky byla pipetována reakční směs o celkovém objemu 17,3 μl podle Tab. 3. Ke kontrolnímu vzorku nebyl přidán templát. Reakce probíhala podle Tab. 4 na str. 29.

Složka	Objem V [μl]
Pufr pro Pfu DNA polymerasu	2
Sterilní voda	13
Primer WRNNUp/WYNNUp (100 μM)	0,1
Primer WRNNDown/WYNNDown (100 μM)	0,1
Nukleotidy dNTPs (10 mM)	2
Templát pUC19-STOP (10 ng/ μl)	0,1

Tab. 3: Složení reakční směsi QuickChange PCR reakce

Fáze	Teplota [°C]	Čas [s]	Počet cyklů
1	95	120	1
2	98	30	32
3	60	30	
4	72	180	
5	72	600	1

Tab. 4: Průběh fází QuickChange PCR reakce

4.5.3 Elektroforéza v agarosovém gelu

Agarosová elektroforéza byla využita k detekci PCR produktů a k izolaci požadovaných fragmentů.

Příprava gelu:

Byl připraven 2% gel smícháním 1 g agarosy v 50 ml TAE nebo TBE elektroforetického pufru. Tato směs byla zahřáta v mikrovlnné troubě tak, aby došlo k úplnému rozpuštění agarosy. Po zchlazení směsi bylo přidáno barvivo GelRed v doporučeném ředění a směs byla nalita na připravenou elektroforetickou desku s tzv. hřebenem. Gel se nechal ztuhnout při pokojové teplotě a poté byl hřeben vyjmut.

Nanášení vzorků:

Gel byl vložen do elektroforetické vany a převrstven elektroforetickým pufrem do vyznačené rysky. Do první jamky bylo aplikováno v rozmezí 2-6 μ l markeru GelPilot 100 bp Plus Ladder. Do ostatních jamek byly aplikovány PCR produkty (obvykle 5 μ l). V případě PCR produktů nemusel být k vzorkům přidáván nanášecí pufr díky barvivu obsaženému v PPP MasterMix v reakční směsi pro PCR reakci (Tab. 1 na str. 27)

Průběh a vyhodnocení:

Elektroforetická vana byla připojena ke zdroji a přiváděné napětí bylo nastaveno na 120 V. Elektroforéza probíhala do doby, než barvivo, putující s čelem, bylo přibližně ve 2/3 gelu. Gel byl poté vyhodnocen pomocí UV transiluminátoru pod světlem o vlnové délce 365 nm. Potřebné fragmenty DNA byly vyříznuty z gelu a dále purifikovány. Pro analytické účely byla použita vlnová délka 312 nm.

4.5.4 Purifikace z gelu

Vyříznutý gel byl přenesen do zvážené sterilní zkumavky a poté byla odečtena hmotnost gelu. DNA byla izolována pomocí komerční sady QIAquick Gel Extraction Kit firmy Qiagen.

Purifikace:

Byly přidány 3 jednotky QG pufru na 1 jednotku gelu (100 mg ~ 100 µl). Zkumavky byly inkubovány v blokovém termostatu při 50°C po dobu 10 minut. Po rozpuštění gelu byla přidána jedna jednotka isopropanolu a zkumavka byla promíchána. Poté byly vzorky přeneseny do sběrných zkumavek s kolonkou a zcentrifugovány ve stolní centrifuze (1 min/16060 g). Proteklá kapalina byla odstraněna ze zkumavky a pro promytí kolonky bylo přidáno 0,75 ml PE pufru. Zkumavky byly zcentrifugovány (1 min/16060 g), proteklá kapalina byla odstraněna. Poté byly zkumavky znovu zcentrifugovány (1 min/16060 g) a proteklá kapalina byla odstraněna. Kolonky byly dány do sterilních 1,5 ml zkumavek a bylo přidáno 50 µl sterilní vody. K vymytí DNA byly zkumavky zcentrifugovány (1 min/16060 g).

4.5.5 Restrikční štěpení

Izolovaná DNA s restrikčními místy byla specificky štěpena pomocí enzymů restrikčních endonukleas.

Štěpení:

Do sterilní zkumavky byla pipetována reakční směs o celkovém objemu 48 µl podle Tab. 5. Reakce probíhala 2 – 4 hodiny při 37°C v inkubátoru. Restrikční štěpení bylo ukončeno snížením teploty.

Složka	Objem V [µl]
Izolovaná DNA	8
Příslušný NEB pufr	5
Restrikční endonukleasa 1	1
Restrikční endonukleasa 2	1
Sterilní voda	33

Tab. 5: Složení reakční směsi restrikčního štěpení

4.5.6 Ligace

Ligací došlo ke spojení fragmentů DNA pomocí DNA ligas. Zde využita T4 DNA ligasa, která je přítomna u bakteriofága T4, je dostupnou ligasou, která umožňuje spojování fragmentů s tupými i s přechýlujícími komplementárními konci.

Ligace:

Do sterilní zkumavky byla pipetována reakční směs o celkovém objemu 20 μ l podle Tab. 6. Jako poslední byla pipetována ligasa. Reakce probíhala 1 hodinu při 20°C v klimatizované místnosti.

Složka	Objem V [μ l]
Pufr pro T4 DNA ligasu	2
Vektor (linearizovaný plasmid)	4
Inzert (vkládaná DNA)	6
Sterilní voda	7
T4 DNA ligasa	1

Tab. 6: Složení reakční směsi ligace

4.5.7 Transformace

Transformací byla plasmidová DNA vnesena do chemicky kompetentních buněk připravených na ÚOCHB AV ČR, v.v.i. z bakteriálních kultur Top10 *E. coli* a Mach1 *E. coli*.

Před začátkem vlastní transformace bylo na již připravené Petriho misky s agarem a ampicilinem (100 μ g/ml) napipetováno 40 μ l zásobního roztoku chromogenního substrátu β -galaktosidasy X-gal (20 mg/ml v DMF), který byl rozetřen sterilní jednorázovou „hocejkou“ (angl. cell spreader). Poté byly plotny dány dnem dolů do inkubátoru o 37°C na 2 hodiny.

Transformace:

Zmražené kompetentní buňky a ligační roztoky byly ponechány na ledu po dobu 5 minut. Poté bylo ke každému ligačnímu roztoku o objemu 20 μ l přidáno 40 μ l kompetentních buněk

a směs byla ponechána na ledu 30 minut. Teplotní šok byl proveden v blokovém termostatu o teplotě 42°C po dobu 90 s a mikrozkušavky byly vráceny na led, kde se ochlazovaly 3 minuty. Nakonec bylo přidáno 100 µl LB média bez antibiotika a po 30 minutách ve 37°C v inkubátoru byla tato bakteriální kultura rozetřena na připravené tuhé médium s antibiotikem a X-gal v Petriho miskách. Ty byly ponechány ve 37°C v inkubátoru přes noc.

4.5.8 Minipreparace plasmidové DNA

Pro přípravu sekvenační metody byla DNA narostlých kolonií izolována a purifikována. DNA byla purifikována pomocí komerční sady QIAprep Spin Miniprep Kit firmy Qiagen.

Kultivace v tekutém médiu:

Jednotlivě narostlé kolonie byly pomocí sterilní špičky přeneseny do 5 ml LB média s ampicilinem (100 µg/ml) a byly inkubovány ve třepačce (220 RPM) při 37°C přes noc. Poté byly bakteriální kultury odstředěny v centrifuze (5 min./3000g) a vzniklý supernatant byl odstraněn.

Vlastní minipreparace:

Vzniklý pelet byl resuspendován ve 250 µl pufru P1 a byl převeden do sterilní zkumavky. K roztoku bylo přidáno 250 µl pufru P2 a zkumavka byla několikrát převrácena k promíchání. Poté bylo přidáno 350 µl pufru N3, roztoky byly znovu promíchány a zcentrifugovány ve stolní centrifuze (10 min/16060 g). Vzniklý supernatant byl přenesen do sterilní zkumavky s kolonkou, vzorky byly zcentrifugovány (1 min/16060 g) a proteklá kapalina byla odstraněna. Poté bylo přidáno 0,5 ml pufru PB, vzorky byly zcentrifugovány (1 min/ 16060 g) a proteklá kapalina byla odstraněna. Kolonka byla nakonec promyta 0,75 ml pufru PE, vzorky byly zcentrifugovány, proteklá kapalina byla odstraněna a vzorky byly ještě jednou stočeny (1 min/16060 g) k úplnému vymytí pufru. Kolonky byly dány do sterilních 1,5 ml zkumavek, bylo přidáno 50 µl sterilní vody, která se nechala jednu minutu působit a poté byly vzorky zcentrifugovány (1 min/16060 g) k vymytí DNA.

Množství získané plasmidové DNA bylo zjišťováno pomocí spektrofotometru NanoDrop, který využívá měření absorbance při 260 nm a zároveň sleduje čistotu vzorku zjišťováním podílu absorbancí při 260 nm a 280 nm.

4.5.9 Sekvenace DNA

Vlastní sekvenace byla provedena firmou GATC Biotech s.r.o.

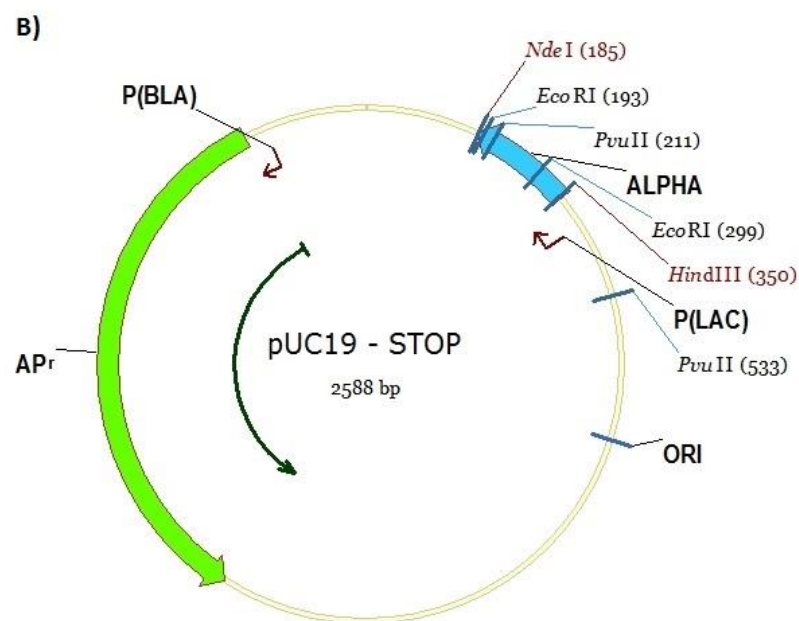
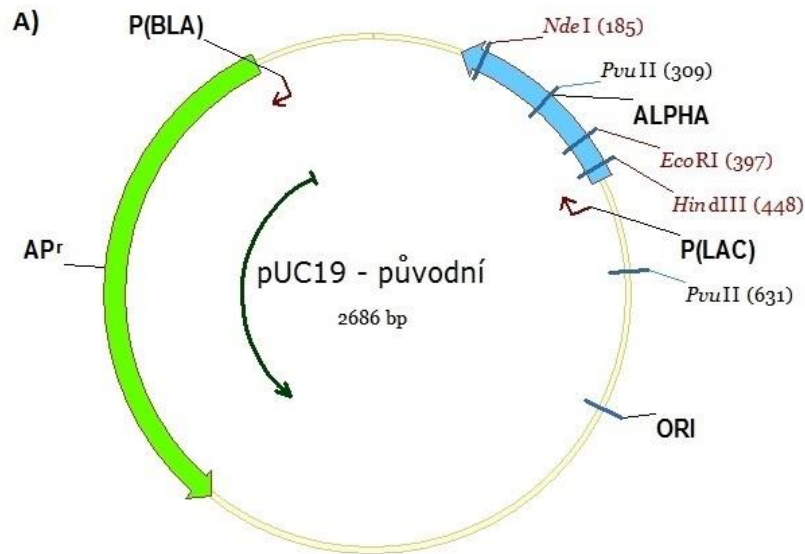
Sekvenace:

Sekvenační směs byla připravována přidáním 2,5 µl primeru M13 rev. (100 µM) k 7,5 µl plasmidové DNA (80 ng/µl).

5 Výsledky

5.1 Ověření funkčnosti zkrácené verze LacZ-alfa komplementačního peptidu

Komerčně dostupný plasmid pUC19 byl na pracovišti ÚOCHB AV ČR, v. v. i. před zahájením této práce geneticky upraven na plasmid s pracovním názvem pUC19 – STOP. Produktem této změny je LacZ-alfa peptid, který je v porovnání s původně kódovaným pUC19 LacZ-alfa peptidem zkrácený na C-konci o 47 zbytků (Obr. 14 na str. 35). Takové zkrácení by mělo zachovat všechny aminokyseliny, které jsou důležité pro α -komplementaci, jak vyplynulo ze strukturních studií této interakce a ve studiích navazujících na tuto práci. Tak se zredukuje studovaný systém na minimální sekvenční prostor.^[56] Dále zde bylo zavedeno nové restriční místo EcoRI, které bude mít důležitou roli při konstrukci knihoven popsané v kapitole 5.2.



C)

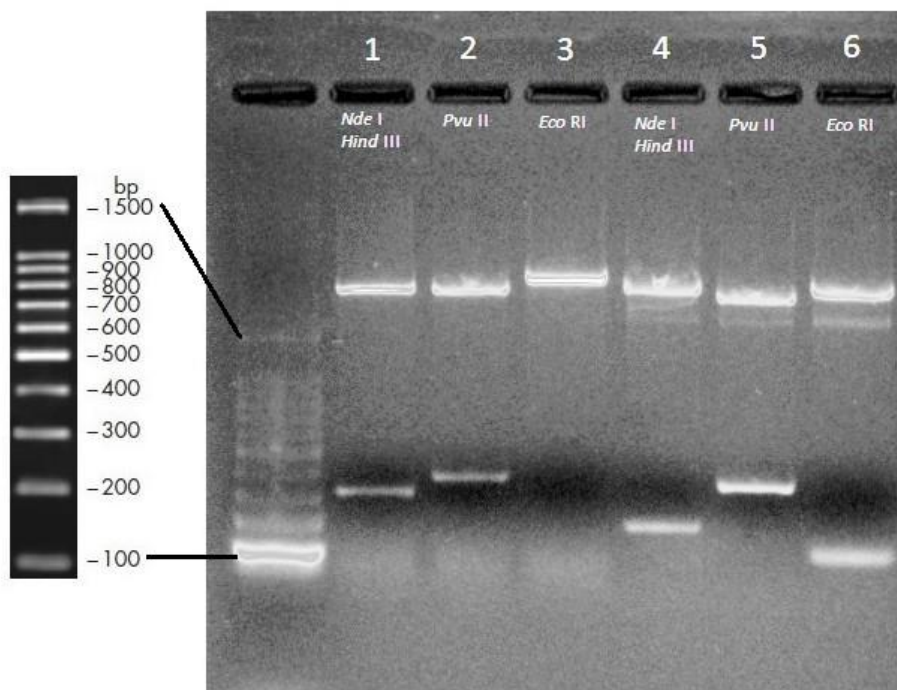
Sekvence pUC19 LacZ-alfa peptidu:

```

MTMITPSLHACRSTLEDPRVPSSNSLA VVLQRRDWENPGVTQLNRLAAH
PPFASWRNSEEARTDRPSQQLRSLNGEWRLMRYFLLTHLCGISHRIWCT
LSTICSDAA
  
```

Obr. 14: **A) Plasmid pUC19, B) pUC19 – STOP** se zkráceným LacZ-alfa peptidem a vneseným restrikčním místem *Eco*RI. ORI značí sekvenci rozpoznávanou DNA polymerasou hostitelské buňky (tedy místo počátku replikace), AP_r značí gen nesoucí rezistenci vůči ampicilinu, P(BLA) značí promotor β-laktamasy, P(LAC) značí promotor α-podjednotky β-galaktosidasy a ALPHA značí LacZ-alfa komplementační peptid. Dále jsou zde vyznačena restrikční místa, která byla důležitá pro další experimenty. Vytvořeno v programu Vector NTI. **C) Sekvence pUC19 LacZ-alfa peptidu:** modře je vyznačena sekvence zkráceného LacZ-alfa peptidu – polylinker je následován úsekem důležitým pro α-komplementaci (podtržen, zbytky 5 – 41)^[56]; žlutě jsou vyznačeny zbytky 61 – 107.

Pro ověření znázorněných umístění restrikčních míst byla provedena agarosová elektroforéza štěpených plasmidů (kap. 4.3.3) a plasmidy byly štěpeny restrikčními endonukleasami (i) *Nde*I a *Hind*III (ii) *Pvu*II a (iii) *Eco*RI. Vyfotografovaný gel je zobrazen na Obr. 15.



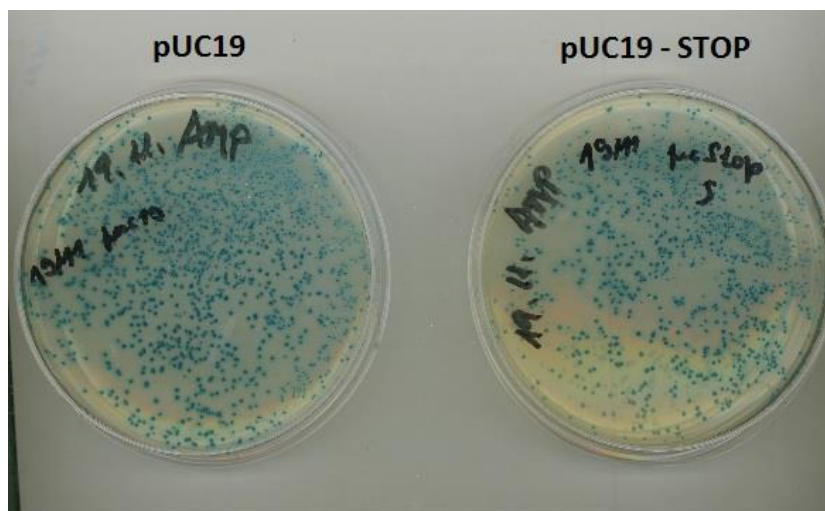
Obr. 15: Agarosová elektroforéza štěpených plasmidů v 1% gelu v TBE pufru. V drahách 1-3 je 14,9 μ l štěpeného původního plasmidu pUC19 (10 ng/ μ l) s přidavkem 3 μ l nanášecího pufru a v drahách 4-6 je 14,9 μ l štěpeného upraveného plasmidu pUC19 – STOP (10 ng/ μ l) s přidavkem 3 μ l nanášecího pufru. Jako marker byly použity 4 μ l GelPilot 100 bp Plus Ladder. Obrázek markeru byl převzat a upraven.^[57]

Z Obr. 15 je patrné, že zavedení druhého restrikčního místa *Eco*RI bylo úspěšné, pUC19 je příslušnou restrikční endonukleasou štěpen pouze jednou (Obr. 15, dráha 3) zatímco pUC19 – STOP je štěpen dvakrát (Obr. 15, dráha 6). Experimentálně pozorované délky fragmentů odpovídají předpokládaným délkám uvedeným v Tab. 7.

	pUC19		pUC19 - STOP	
<i>Nde</i> I, <i>Hind</i> III	2423 bp	263 bp	2423 bp	165 bp
<i>Pvu</i> II	2364 bp	322 bp	2266 bp	322 bp
<i>Eco</i> RI	2686 bp		2482 bp	106 bp

Tab. 7: Tabulka s předpokládanými délkami fragmentů plasmidů pUC19 a pUC19 – STOP.

Dále byly oba plasmidy transformovány (kap. 4.3.7) do buněk *E. coli* kódujících LacZ-omega peptid (kmeny Top10 a Mach1) a kultivovány na tuhém médiu (kap. 4.3.8). Z Obr. 16 vyplývá, že oba plasmidy obsahují funkční LacZ-alfa peptid a klony produkují aktivní β -galaktosidasu, která štěpí X-gal za vzniku modře zbarvených kolonií.



Obr. 16: Kolonie produkující aktivní β -galaktosidasu.

5.2 Konstrukce knihoven - zavedení mutace

Tento experiment sloužil k ověření metodiky pozdějšího obohacování LacZ-alfa peptidu o předpokládané rané aminokyseliny.

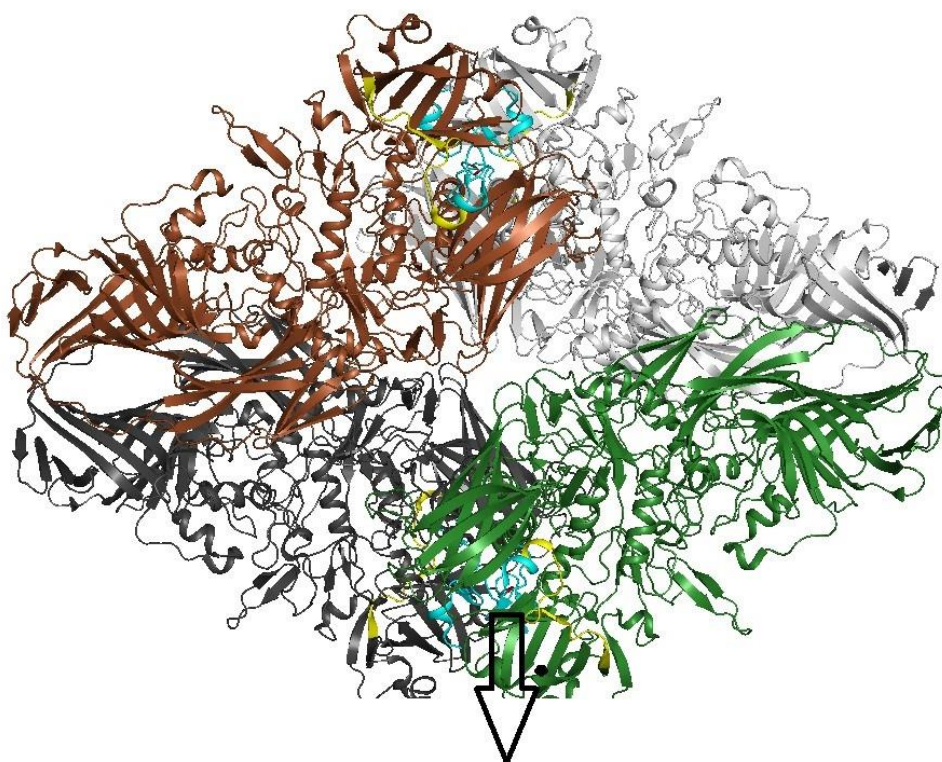
Vycházeli jsme z tabulky univerzálního genetického kódu s rozděleními aminokyselinami podle jejich “stáří“, kde je její dolní část obohacena aminokyselinami považovanými za rané a horní část je obohacena aminokyselinami považovanými za evolučně novější (Obr. 6 na str. 11).

Výchozí sekvence zkráceného LacZ-alfa peptidu je zobrazena na Obr. 17.

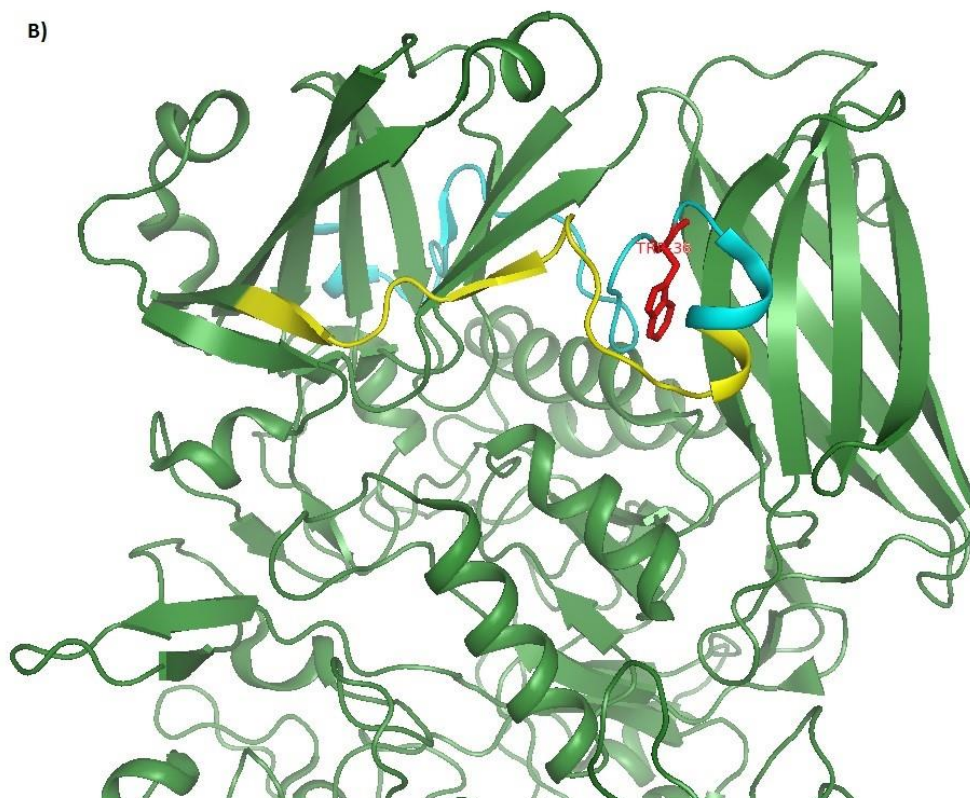
MTMITPSLHACRSTLEDPRVPSSNSLAVVLQ
RRDWENPGVTQLNRLAAHPPFAS(W)RNSEE

Obr. 17: **Sekvence** zkráceného LacZ-alfa peptidu, kde barevná znázornění odpovídají rozdělení aminokyselin podle Trifonova (Obr. 6 na str.11). Pro mutagenézi byl vybrán zakroužkovaný tryptofan na C-konci peptidu.

A)



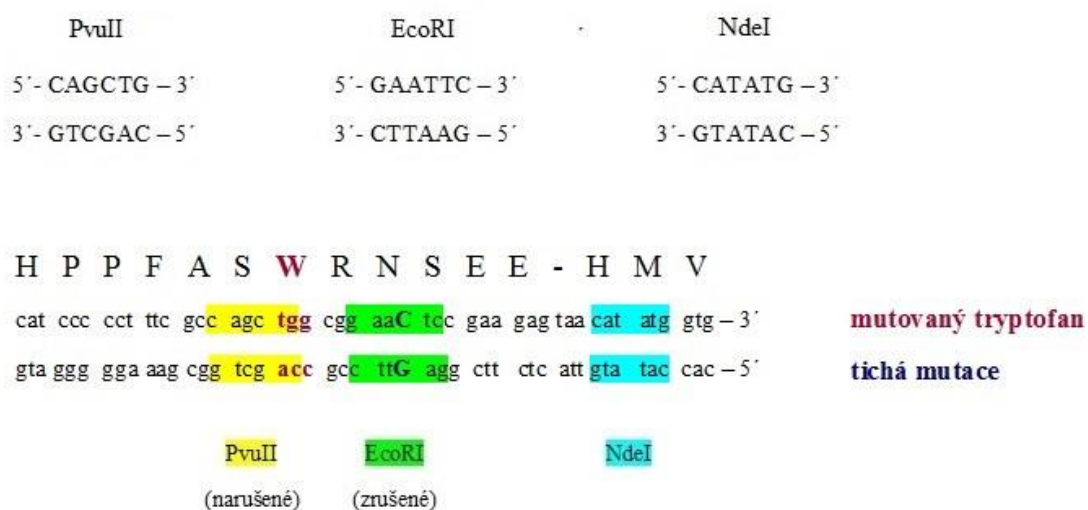
B)



Obr. 18: **A)** β -galaktosidasa, kde jsou 4 podjednotky barevně rozlišeny, každá podjednotka má žlutě vyznačený lacZ-alfa peptid, modře vyznačenou sledovanou sekvenci a červeně je vyznačen mutovaný tryptofan **B)** přiblížená sekvence jedné podjednotky s červeně označeným **mutovaným tryptofanem Trp36**. Vytvořeno v programu PyMol.

Byly navrženy primery, které by mutovaly daný tryptofan v součtu na všech 20 možných proteinogenních aminokyselin a to ve 2 různých knihovnách podle jejich "stáří". Konkrétně se tedy jednalo o mutaci kodonu tryptofanu na A/GNN (primery WYNNDown a WYNNUp), kdy by vznikaly s větší četností rané aminokyseliny (z dolní části tabulky podle Trifonova, Obr. 6 na str. 11) a T/CNN, kódující ve větší míře evolučně novější aminokyseliny (z horní části tabulky, Obr. 6 na str. 11). (kap. 4.4.1 a 4.4.2).

Do těchto primerů byla zároveň zavedená tichá mutace, která ruší původně zavedené restrikční místo EcoRI. To umožní rozlišit klony nesoucí tryptofan vnesený mutagenézí od původního plasmidu. Mutací tryptofanu na zbylých 63 kodonů dochází k narušení restrikčního místa PvuII, které zde bude dále přítomno pouze v případě vzniku tryptofanu, cysteinu nebo Stop kodonu TGA (Obr. 19).



Obr. 19: **C-koncová část** zkráceného LacZ-alfa peptidu s vyznačenými změnami restrikčních míst při mutagenézi tryptofanu.

Vzniklý klon tedy bude moci mít 3 různá uspořádání restrikčních míst (Tab. 8).

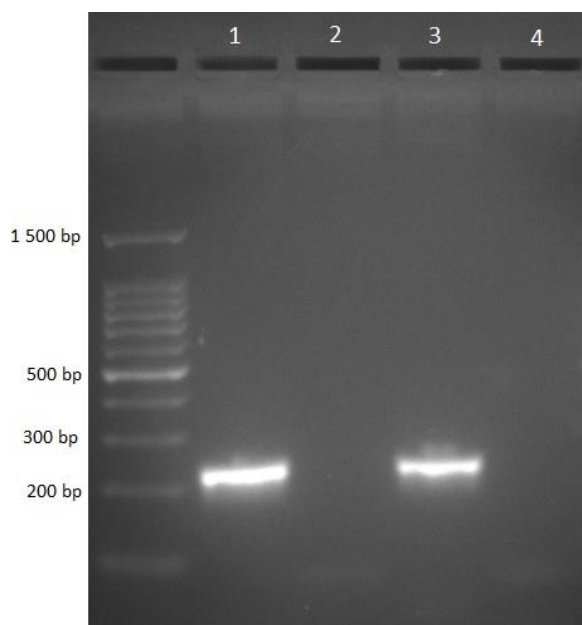
PvuII	EcoRI	Varianta plasmidu
ANO	ANO	Původní plasmid s tryptofanem
ANO	NE	Klon s tryptofanem, cysteinem nebo Stop kodonem TGA
NE	NE	Klon s ostatními možnostmi kodonů

Tab. 8: Varianty plasmidů na základě přítomnosti restrikčních míst.

Mutageneze byla provedena dvěma metodami:

- (i) mutageneze s použitím dvojice primerů vymezujících amplifikovaný úsek (kap. 4.3.1)

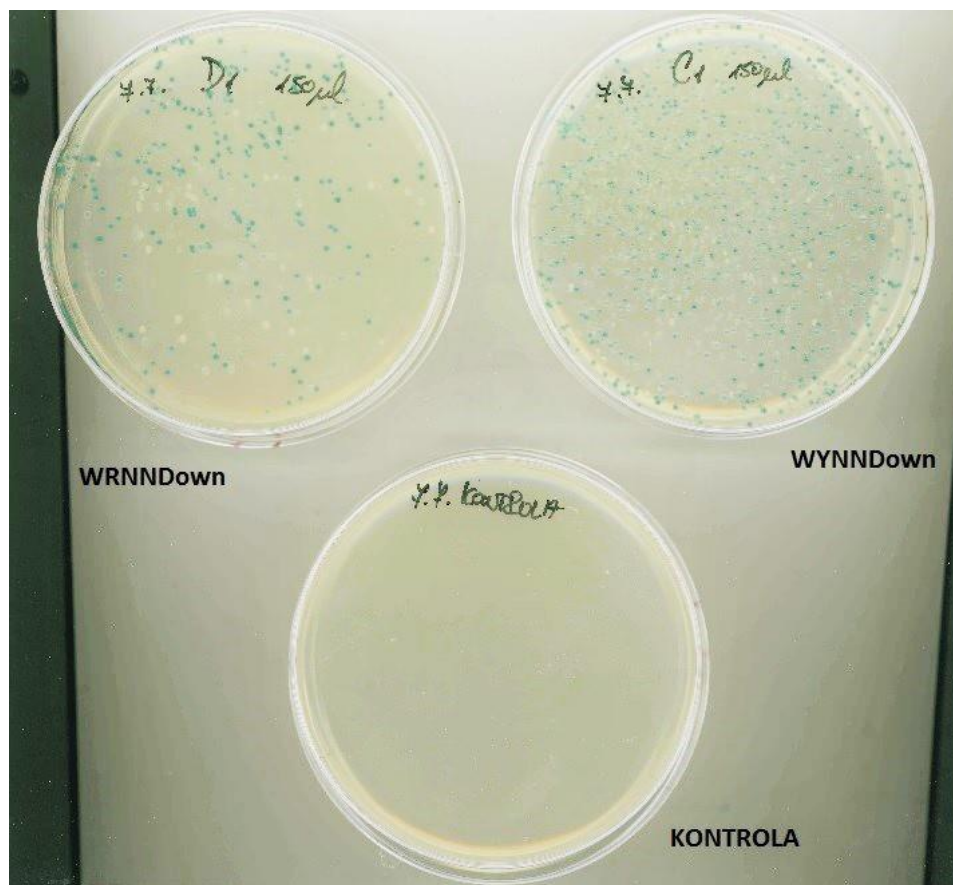
Detekce PCR produktů byla provedena pomocí agarosové elektroforézy (kap. 4.3.3) ve 2% agarosovém gelu. Vyfotografovaný gel je zobrazen na Obr. 20.



Obr. 20: Detekce PCR produktů v 2% agarosovém gelu v TAE pufru. Jako templát byl použit plasmid pUC19 – STOP. V dráze 1 je PCR produkt získaný s použitím primerů M13 rev. a WYNNDown, v dráze 2 je odpovídající negativní kontrola, v dráze 3 je PCR produkt získaný s použitím primerů M13 rev. a WRNNDDown, v dráze 4 je odpovídající negativní kontrola. Do všech 4 drah bylo napipetováno 5 μ l vzorku. Jako marker byly použity 2 μ l GelPilot 100 bp Plus Ladder.

PCR produkty z dráhy 1 a 3 na Obr. 20 byly poté vyříznuty a purifikovány z gelu (kap. 4.3.4). Dále následovalo restriční štěpení (NdeI a HindIII) (kap. 4.4.5), ligace (kap. 4.4.6), transformace do vhodného kmene *E. coli* (kap. 4.4.7) a kultivace na tuhém médiu (kap. 4.4.8).

Příprava obou knihoven metodou místně specifické mutageneze s použitím dvojice primerů vymezujících amplifikovaný úsek byla úspěšná. Z obou variant PCR produktů byla získána knihovna mutantních plasmidů, které daly vznik koloniím. Tyto kolonie byly jednak bílé a jednak byly v různých odstínech modré, z čehož se dalo usoudit, že jsou zde přítomny různé varianty LacZ-alfa peptidu (Obr. 21 na str. 41).



Obr. 21: **Kolonie** obou knihoven a negativní kontrola.

- (ii) mutageneze řízená dvojicí komplementárních primerů (kap. 4.3.2)
Zavedení mutací mutagenezí touto metodou nebylo úspěšné, po transformaci jsme nezískali žádné kolonie.

5.3 Vyhodnocení

U vzniklých kolonií byla nejdříve sledována jejich barva (fenotyp) a poté byla zjišťována sekvence (genotyp) vybraných plasmidů získaných z individuálních kolonií.

5.3.1 Fenotypizace knihoven

U obou knihoven byly pozorovány bílé a modré kolonie, kdy modrá barva byla v několika různých odstínech. Různorodost modrého zabarvení byla zjednodušena rozdělením na modré a světle modré kolonie.

Byly sledovány počty kolonií jednotlivých knihoven a výsledky jsou uvedeny v Tab. 9 a Tab. 10.

Knihovna obohacená ranými aminokyselinami				
Barva kolonie	1. sada*	2. sada*	3. sada [#]	celkem
modrá	140	151	1016	1307
světle modrá	117	63	536	716
bílá	9	2	664	675

Tab. 9: Počet kolonií v knihovně obohacené ranými aminokyselinami, kde * značí Top10 *E. coli* a [#] značí Mach1 *E. coli*.

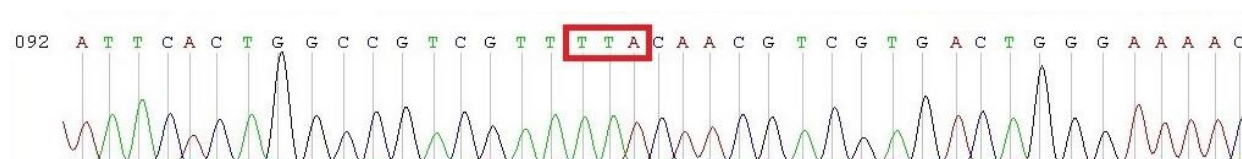
Knihovna obohacená evolučně novějšími aminokyselinami				
Barva kolonie	1. sada*	2. sada*	3. sada [#]	celkem
modrá	46	71	412	529
světle modrá	3	24	631	658
bílá	3	3	455	461

Tab. 10: Počet kolonií v knihovně obohacené evolučně novějšími aminokyselinami, kde * značí Top10 *E. coli* a [#] značí Mach1 *E. coli*.

Podle statistických výpočtů v kapitole 2.3.1.1, kdy pro 99% pravděpodobnost získání kompletní knihovny vychází nutný počet získaných kolonií 258, můžeme usoudit, že získaný počet kolonií pokryl všechny možné varianty mutovaného LacZ-alfa peptidu.

5.3.2 Genotypizace knihoven

Vybrané kolonie poté byly kultivovány v tekutém médiu a následně byla provedena minipreparace DNA (kap. 4.4.9). Takto připravené vzorky byly osekvenovány (kap. 4.4.10). Příklad části výstupu ze sekvenace je na Obr. 22.



Obr. 22: Příklad části výstupních dat **sekvenace**, kde je triplet kódující leucin v pozici původního tryptofanu (červený obdélník). Sekvenace byla provedena firmou GATC Biotech s.r.o.

5.3.2.1 Knihovna obohacená ranými aminokyselinami

Z knihovny obohacené ranými aminokyselinami (dolní část tabulky univerzálního genetického kódu) bylo genotypicky charakterizováno 43 kolonií a byly zde nalezeny všechny varianty možných aminokyselin (Obr. 23).

	T	C	A	G	
A	Ile	Thr	Asn	Ser	T
	Met		Lys	Arg	C
					A
				G	
G	Val	Ala	Asp	Gly	T
			Glu		C
					A
				G	

barva kolonie:

modrá

světle modrá

Obr. 23: Tabulka raných aminokyselin s **vyznačeným fenotypem** jednotlivých identifikovaných variant.

	T	C	A	G	
A	Ile (6x)	Thr	Asn (3x)	Ser	T
	Ile (3x)	Thr (1x)	Asn (1x)	Ser (1x)	C
	Ile (2x)	Thr (1x)	Lys (1x)	Arg (2x)	A
	Met (2x)	Thr (2x)	Lys (1x)	Arg	G
G	Val	Ala (1x)	Asp (6x)	Gly (1x)	T
	Val	Ala (1x)	Asp (2x)	Gly	C
	Val	Ala	Glu (2x)	Gly	A
	Val (1x)	Ala (2x)	Glu (1x)	Gly	G

barva kolonie:

modrá

světle modrá

nenalezeno

Obr. 24: Tabulka jednotlivých **variant tripletů**, kde nalezené varianty jsou rozděleny podle fenotypu. Číslo v závorce udává počet nalezených klonů dané varianty.

Z Obr. 24 vyplývá, že nebyly nalezeny všechny varianty možných tripletů v této knihovně, což ale nebylo účelem tohoto experimentu.

Bylo získáno 22 z 32 možných tripletů (69%), kdy toto množství stačilo na pokrytí všech možných variant aminokyselin.

Knihovna obohacená ranými aminokyselinami			
báze	1. pozice	2. pozice	3. pozice
T	-	14	17
C	-	8	9
A	26	17	8
G	17	4	9

Tab. 11: Počet bazí v jednotlivých pozicích mutovaného tripletu.

Z Tab. 11 vyplývá, že nebyl zachován předpokládaný poměr bazí A:G 1:1, stejně tak nebyly rovnoměrně rozděleny poměry bazí ve zbylých dvou pozicích tripletu.

5.3.2.2 Knihovna obohacená evolučně novějšími aminokyselinami

Z knihovny obohacené evolučně novějšími aminokyselinami (horní část tabulky univerzálního genetického kódu) bylo genotypicky charakterizováno 113 kolonií a byly zde nalezeny všechny varianty možných aminokyselin (Obr. 25).

	T	C	A	G	
T	Phe	Ser	Tyr	Cys	T
	Leu		Stop	Trp	C
					A
					G
C	Leu	Pro	His	Arg	T
			Gln		C
					A
					G

barva kolonie:

modrá

světle modrá

bílá

Obr. 25: Tabulka evolučně novějších aminokyselin s vyznačeným fenotypem jednotlivých identifikovaných variant.

	T	C	A	G	
T	Phe (13x)	Ser (2x)	Tyr (11x)	Cys	T
	Phe (6x)	Ser (3x)	Tyr (5x)	Cys (2x)	C
	Leu (12x)	Ser (2x)	Stop (5x)	Stop (1x)	A
	Leu (2x)	Ser	Stop (2x)	Trp (1x)	G
C	Leu (3x)	Pro (1x)	His (6x)	Arg (1x)	T
	Leu (1x)	Pro (2x)	His (4x)	Arg (2x)	C
	Leu (8x)	Pro (2x)	Gln (6x)	Arg (1x)	A
	Leu (3x)	Pro (1x)	Gln (2x)	Arg (3x)	G

barva kolonie:

modrá

světle modrá

bílá

nenalezeno

Obr. 26: Tabulka jednotlivých **variant tripletů**, kde nalezené varianty jsou rozděleny podle fenotypu. Číslo v závorce udává počet nalezených klonů dané varianty.

Z Obr. 26 vyplývá, že nebyly nalezeny všechny varianty možných tripletů v této knihovně. Avšak, jak již bylo zmíněno výše, to nebylo účelem daného experimentu.

Bylo získáno 30 z 32 možných tripletů (94%), kdy toto množství pokrylo všechny možné varianty aminokyselin.

Knihovna obohacená evolučně novějšími aminokyselinami			
báze	1. pozice	2. pozice	3. pozice
T	67	48	37
C	46	13	25
A	-	41	37
G	-	11	14

Tab. 12: Počet bazí v jednotlivých pozicích mutovaného tripletu.

Z Tab. 12 vyplývá, že nebyl zachován předpokládaný poměr bazí T:C 1:1, stejně tak nebyly rovnoměrně rozděleny poměry bazí ve zbylých dvou pozicích tripletu.

6 Diskuse

Dnes je genetický kód převážně považován za výsledek evoluce raného genetického kódu, který obsahoval méně aminokyselin než ten současný. Otázkou zůstává, zda mohly být prvotní proteiny složené z redukovaného počtu aminokyselin funkční. Další otázkou je jaký byl klíč k výběru současných 20ti proteinogenních aminokyselin z rozsáhlého souboru dostupných aminokyselin. Jednou z možností jak se přiblížit k odpovědím na tyto otázky je studium současných proteinů, ve kterých jsou aminokyseliny považované za evolučně novější nahrazeny aminokyselinami považovanými za rané.

Hlavním cílem předkládané práce bylo připravit metodiku mutagenese zkráceného LacZ-alfa peptidu za účelem pozdějšího studia jeho variant obohacených ranými aminokyselinami.

Pro mutagenesi jsme vytipovali C-koncový tryptofan, který je pravděpodobně jednou z evolučně nejmladších aminokyselin.^[24] Dalším předpokladem pro jeho výběr byla skutečnost, že jako jedna ze dvou proteinogenních aminokyselin nemá degenerovaný kodon. A nakonec jeho pozice v LacZ-alfa peptidu, která se zdá být strukturálně méně významná (Obr. 18 na str. 38), kdy se daný tryptofan nachází v místě exponovaném do rozpouštědla a nemá významnější interakce s ostatními aminokyselinami. Při konstrukci knihovny místně specifických mutantů LacZ-alfa peptidu jsme vycházeli z rozdělení tabulky univerzálního genetického kódu podle Trifonova (Obr. 6 na str. 11), tedy její horní poloviny obohacené evolučně novějšími aminokyselinami a dolní poloviny obohacené ranými aminokyselinami.

Příprava obou knihoven metodou mutagenese řízené dvojicí komplementárních primerů se neosvědčila. Pravděpodobným důvodem nefunkčnosti této metody je použití Pfu polymerasy, která má díky své 3'→5' exonukleasové aktivitě schopnost opravy špatně přiřazených bazí (angl. proofreading). Díky ní ovšem může docházet k degradaci DNA ve vzorku, která byla v našem případě podpořena přítomností nespárovaných tripletů.

Příprava obou knihoven metodou místně specifické mutagenese s použitím dvojice primerů vymezujících amplifikovaný úsek byla úspěšná. Po úspěšné ligaci a transformaci PCR produktů jsme získali dostatečné množství kolonií s různým fenotypem. Většina kolonií měla různé odstíny modré barvy, z čehož jsme mohli usuzovat přítomnost

mutovaných variant. Dalším důvodem rozdílnosti v odstínech mohlo být nerovnoměrné rozetření chromogenního substrátu X-gal na agarové misce. Pro budoucí experimenty je tedy důležité důsledné rozetření tohoto substrátu nebo by další možností bylo přidání X-gal do tekutého agaru před nalitím na misky. Přestože jsme se pokusili tyto odstíny rozdělit na modrou a světle modrou barvu, byl tento výběr modrých variant poněkud subjektivní. Na druhou stranu množství různých odstínů umožňovalo jednoduší výběr rozdílných variant mutací, a to spíše v knihovně obohacené ranými aminokyselinami. Díky této vlastnosti bylo pro získání všech variant aminokyselin potřeba menší počet odebraných klonů k sekvenaci. Zatímco u knihovny obohacené evolučně novějšími aminokyselinami jsme pozorovali menší rozdíly v odstínech modré, a také proto bylo získání co možná největšího množství možných variant obtížnější. Z této knihovny byla získána varianta s nově vzniklým tryptofanem, u kterého byl dopředu znám fenotyp. Tato skutečnost může být považována za pozitivní kontrolu.

Z genotypové analýzy vyplynulo, že Trp36 je možné zaměnit za jakoukoliv jinou aminokyselinu (kromě Stop kodonů) při zachování α -komplementace. To nebyl příliš překvapivý výsledek vzhledem k tomu, že tato pozice byla vybrána i na základě malé strukturní role. Světle modré kolonie byly pozorovány při záměně za Ala, Gly, Glu, Asp a Pro. Statisticky nevýhodná záměna Trp za tyto aminokyseliny je popsána v substitučních maticích a je opodstatněné se domnívat, že světlé zabarvení kolonií tedy mohlo značit strukturně-vazebně nevýhodné záměny.^[58]

Z genotypování také vyplynulo, že zastoupení jednotlivých bazí v mutované pozici neodpovídá předpokládaným poměrům, tedy na první pozici 1:1 a na dalších dvou pozicích 1:1:1:1. Například pro knihovnu obohacenou ranými aminokyselinami byl poměr bazí na první pozici 26:17, zatímco pro knihovnu obohacenou evolučně novějšími aminokyselinami byl poměr bazí na první pozici 67:46. To může poukazovat na syntetický artefakt vzniklý při syntéze oligonukleotidů. Účinnost připojování (z angl. coupling) jednotlivých bazí byla diskutována přes poradenský portál firmy GENERI BIOTECH s.r.o., u které byly primery objednány. Dle jejich odpovědi je v pokynech dodaných k přístroji uvedeno pořadí účinnosti připojování následující: T, G, C, A (tedy nejlépe se váže báze T). Toto však je jen obecné pravidlo. Poměr bazí v konkrétních degenerovaných oligonukleotidech se tedy může lišit a tento poznatek by měl být zohledněn v navazujících studiích. Sledování zastoupení bazí v mutovaných pozicích nebylo prvotním záměrem této

práce, a proto byly také kolonie vybírány se zřetelem na různé zastoupení fenotypu, a když byly například nalezeny všechny varianty aminokyselin u knihovny obohacené ranými aminokyselinami, byla genotypizace ukončena. Z toho důvodu bylo u této knihovny analyzováno jen 43 kolonií, zatímco u knihovny obohacené evolučně novějšími aminokyselinami bylo analyzováno 113 kolonií. Podle statistických výpočtů (kap. 2.3.1.1) bylo zjištěno, že pro knihovnu, která pokryje 95% variant, je třeba analyzovat 96 klonů. Pro důkladnou analýzu výskytu jednotlivých kodonů tedy nebyl u jedné z knihoven tento výpočet následován a pro vyvození závěrů by bylo třeba tuto analýzu dále rozšířit.

Budoucí studie námi otestované sekvence LacZ-alfa peptidu bude mít za cíl randomizovat celkem 13 pozic, které jsou v její současné sekvenci obsazeny evolučně novějšími aminokyselinami. Randomizace bude v této další studii probíhat jen za rané aminokyseliny (a v některých případech dokonce jen za jejich podmnožinu), ale i přesto budou tvořeny knihovny s velkým množstvím variant. Praktické poznatky o statistickém pokrytí variant z této práce jsou tedy pro další postup velmi přínosné.

7 Závěr

Bylo ověřeno, že zkrácený pUC19 LacZ-alfa peptid (obsahující polylinker a zbytky 5 – 41) je funkční pro účely α -komplementace a modro/bílé selekce.

Metodou místně specifické mutageneze byly připraveny dvě knihovny variant v místě Trp36 obohacené o (i) rané a (ii) evolučně novější aminokyseliny. Pro knihovnu obohacenou ranými aminokyselinami bylo fenotypicky charakterizováno 2698 variant a z nich dále 43 variant bylo charakterizováno genotypicky. Pro knihovnu obohacenou evolučně novějšími aminokyselinami bylo fenotypicky charakterizováno 1648 variant a z nich dále 113 variant bylo charakterizováno genotypicky.

U obou knihoven byly pozorovány různé fenotypy (bílé a modré kolonie, kde modré měly různé odstíny) a Trp36 byl při zachování α -komplementace úspěšně zaměněn všemi 20ti proteinogenními aminokyselinami. Zabarvení kolonií nasvědčovalo tomu, že záměna za některé z raných aminokyselin (Ala, Gly, Asp, Glu, Pro) byla v porovnání s ostatními méně vhodná pro zachování α -komplementace.

Z genotypové analýzy získaných variant vyplynulo, že oligonukleotidy s degenerovanými pozicemi, které byly použity pro místně specifické mutageneze, zřejmě neobsahovaly rovnoměrné poměry jednotlivých bazí a tak jednotlivé kodony nevznikaly se shodnou pravděpodobností. To je přínosné zjištění při zavádění této metodiky pro další studie.

Předkládaná práce potvrdila vhodnost použití LacZ-alfa peptidu pro další mikroevoluční studie s použitím α -komplementace. Jako vhodné zavedení mutace se ukázala mutageneze s použitím dvojice primerů vymezujících amplifikovaný úsek. Naopak jako nevhodné zavedení mutace se ukázala mutageneze řízená dvojicí komplementárních primerů.

8 Seznam použité literatury

- [1] Crick, F.: Central dogma of molecular biology. *Nature* **227**, 561–563 (1970).
- [2] Freeland, S.: "Terrestrial" amino acids and their evolution, v knize *Amino Acids, peptides and proteins in organic chemistry* **1**, (Hughes, A.B. ed.) Weinheim, Germany, str. 43–75 (2010).
- [3] Szathmary, E.: The origin of the genetic code: amino acids as cofactors in an RNA world. *Trends Genet* **15**, 223–229 (1999).
- [4] Rosypal, S.: *Úvod do molekulární biologie (první díl)*. Stanislav Rosypal, Brno (2006).
- [5] Osawa, S., Jukes, T. H., Watanabe, K. & Muto, A.: Recent evidence for evolution of the genetic code. *Microbiol. Rev.* **56**, 229–64 (1992).
- [6] Lukashenko, N. P.: Expanding genetic code: Amino acids 21 and 22, selenocysteine and pyrrolysine. *Russ. J. Genet.* **46**, 899–916 (2010).
- [7] Bada, J. L.: How life began on Earth: A status report. *Earth Planet. Sci. Lett.* **226**, 1–15 (2004).
- [8] Chyba, C. & Sagan, C.: Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life. *Nature* **355**, 125–132 (1992).
- [9] Cleaves, H. J.: The origin of the biologically coded amino acids. *J. Theor. Biol.* **263**, 490–498 (2010).
- [10] Pudritz, R., Higgs, P. & Stone, J.: *Planetary Systems and the Origins of Life*. Cambridge University Press, New York (2007).
- [11] Rauchfuss, H.: *Chemical Evolution and the Origin of Life*. Springer Berlin Heidelberg (2008).
- [12] Burton, A. S., Stern, J. C., Elsila, J. E., Glavin, D. P. & Dworkin, J. P.: Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites. *Chem. Soc. Rev.* **41**, 5459 (2012).
- [13] Zaia, D. A. M., Zaia, C. T. B. V. & De Santana, H.: Which amino acids should be used in prebiotic chemistry studies? *Orig. Life Evol. Biosph.* **38**, 469–488 (2008).
- [14] Miller, S. L.: A production of amino acids under possible primitive earth conditions. *Science* **117**, 528–529 (1953).

- [15] Oparin, A. I.: The origin of life. *Nord. Med.* **65**, 693–697 (1961).
- [16] Ring, D., Wolman, Y., Friedmann, N. & Miller, S. L.: Prebiotic synthesis of hydrophobic and protein amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **69**, 765–768 (1972).
- [17] Bada, J. L.: New insights into prebiotic chemistry from Stanley Miller’s spark discharge experiments. *Chem. Soc. Rev.* **42**, 2186–96 (2013).
- [18] Johnson, A. P. *et al.*: The Miller volcanic spark discharge experiment. *Science* **322**, 404 (2008).
- [19] Higgs, P. G. & Pudritz, R. E.: A thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code. *Astrobiology* **9**, 483–490 (2009).
- [20] Brooks, D. J., Fresco, J. R., Lesk, A. M. & Singh, M.: Evolution of amino acid frequencies in proteins over deep time: inferred order of introduction of amino acids into the genetic code. *Mol. Biol. Evol.* **19**, 1645–1655 (2002).
- [21] Cunchillos, C. & Lecointre, G.: Integrating the universal metabolism into a phylogenetic analysis. *Mol. Biol. Evol.* **22**, 1–11 (2005).
- [22] Trifonov, E. N.: Consensus temporal order of amino acids and evolution of the triplet code. *Gene* **261**, 139–151 (2000).
- [23] Rost, B.: Did evolution leap to create the protein universe? *Curr. Opin. Struct. Biol.* **12**, 409–416 (2002).
- [24] Trifonov, E. N.: The triplet code from first principles. *J. Biomol. Struct. Dyn.* **22**, 1–11 (2004).
- [25] Yanagawa, H.: Exploration of the origin and evolution of globular proteins by mRNA display. *Biochemistry* **52**, 3841–3851 (2013).
- [26] McDonald, G. D. & Storrie-Lombardi, M. C.: Biochemical constraints in a protobiotic earth devoid of basic amino acids: the „BAA(-) world". *Astrobiology* **10**, 989–1000 (2010).
- [27] Weber, A. L. & Miller, S. L.: Reasons for the occurrence of the twenty coded protein amino acids. *J. Mol. Evol.* **17**, 273–284 (1981).
- [28] Doi, N., Kakukawa, K., Oishi, Y. & Yanagawa, H.: High solubility of random-sequence proteins consisting of five kinds of primitive amino acids. *Protein Eng. Des. Sel.* **18**, 279–284 (2005).
- [29] Tanaka, J., Doi, N., Takashima, H. & Yanagawa, H.: Comparative characterization of random-sequence proteins consisting of 5, 12, and 20 kinds of amino acids. *Protein Sci.* **19**, 786–795 (2010).

- [30] Tanaka, J., Yanagawa, H. & Doi, N.: Comparison of the frequency of functional SH3 domains with different limited sets of amino acids using mRNA display. *PLoS One* **6**, (2011).
- [31] Murphy, L. R., Wallqvist, A. & Levy, R. M.: Simplified amino acid alphabets for protein fold recognition and implications for folding. *Protein Eng.* **13**, 149–152 (2000).
- [32] Riddle, D. S. *et al.*: Functional rapidly folding proteins from simplified amino acid sequences. *Nat. Struct. Biol.* **4**, 805–809 (1997).
- [33] Longo, L. M., Lee, J. & Blaber, M.: Simplified protein design biased for prebiotic amino acids yields a foldable, halophilic protein. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 2135–9 (2013).
- [34] Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M. & Hecht, M. H.: Protein design by binary patterning of polar and nonpolar amino acids. *Science* **262**, (1993).
- [35] Go, A., Kim, S., Baum, J. & Hecht, M. H.: Structure and dynamics of *de novo* proteins from a designed superfamily of 4-helix bundles. *Protein Sci.* **17**, 821–832 (2008).
- [36] Patel, S. C., Bradley, L. H., Jinadasa, S. P. & Hecht, M. H.: Cofactor binding and enzymatic activity in an unevolved superfamily of *de novo* designed 4-helix bundle proteins. *Protein Sci.* **18**, 1388–1400 (2009).
- [37] Walter, K. U., Vamvaca, K. & Hilvert, D.: An active enzyme constructed from a 9-amino acid alphabet. *J. Biol. Chem.* **280**, 37742–37746 (2005).
- [38] Jumawid, M. T., Takahashi, T., Yamazaki, T., Ashigai, H. & Mihara, H.: Selection and structural analysis of *de novo* proteins from an $\alpha\beta\beta$ genetic library. *Protein Sci.* **18**, 384–398 (2009).
- [39] Akanuma, S., Kigawa, T. & Yokoyama, S.: Combinatorial mutagenesis to restrict amino acid usage in an enzyme to a reduced set. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 13549–13553 (2002).
- [40] Brown, B. M. & Sauer, R. T.: Tolerance of Arc repressor to multiple-alanine substitutions. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 1983–1988 (1999).
- [41] Babajide, A., Hofacker, I. L., Sippl, M. J. & Stadler, P. F.: Neutral networks in protein space: a computational study based on knowledge-based potentials of mean force. *Fold. Des.* **2**, 261–269 (1997).
- [42] Packer, M. S. & Liu, D. R.: Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379–394 (2015).
- [43] Golynskiy, M. V., Haugner, J. C., Morelli, A., Morrone, D. & Seelig, B.: *In vitro* evolution of enzymes. *Methods Mol. Biol.* **978**, 73–92 (2013).

- [44] Rinnová, M.: Knihovny struktur a chemická syntéza. *Chem. List.* **4**, (1994).
- [45] Patrick, W. M., Firth, A. E. & Blackburn, J. M.: User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Eng.* **16**, 451–457 (2003).
- [46] Müller-Hill, B.: *The lac Operon*. Walter de Gruyter, Berlin, New York (1996).
<https://books.google.com/books?id=ppRmC9-a6JQC&pgis=1> (12.8. 2015)
- [47] Juers, D. H., Matthews, B. W. & Huber, R. E. LacZ β -galactosidase: Structure and function of an enzyme of historical and molecular biological importance. *Protein Science* **21**, 1792–1807 (2012).
- [48] Jacobson, R. H., Zhang, X.-J., Dubose, R. F. & Matthews, B. W.: Three-dimensional structure of β -galactosidase from *E. coli*. *Nature* **369**, 761–766 (1994).
- [49] Matthews, B. W.: The structure of *E. coli* β -galactosidase. *C. R. Biol.* **328**, 549–556 (2005).
- [50] Shuman, H. A. & Silhavy, T. J.: The art and design of genetic screens: *Escherichia coli*. *Nat. Rev. Genet.* **4**, 419–431 (2003).
- [51] Mogalisetti, P. & Walt, D. R.: Stoichiometry of the α -complementation reaction of *Escherichia coli* β -galactosidase as revealed through single-molecule studies. *Biochemistry* **54**, 1583–8 (2015).
- [52] Ruml, T., Rumlová, M. & Pačes, V.: *Genové inženýrství*. VŠCHT Praha (2012).
- [53] <http://www.sigmaaldrich.com/catalog/product/sigma/d3404?lang=en®ion=CZ> (12.8. 2015)
- [54] <https://www.neb.com/products/n3041-puc19-vector> (12.8. 2015)
- [55] Berg, J. M., Tymoczko, J. L. & Stryer, L.: *Biochemistry*. W. H. Freeman, str. 896 (2006).
- [56] Juers, D. H. *et al.*: High resolution refinement of beta-galactosidase in a new crystal form reveals multiple metal-binding sites and provides a structural basis for alpha-complementation. *Protein Sci.* **9**, 1685–1699 (2000).
- [57] <https://www.qiagen.com/cz/shop/lab-essentials-and-accessories/gelpilot-dna-molecular-weight-markers/#productdetails> (12.8. 2015)
- [58] Overington, J., Donnelly, D., Johnson, M. S., Šali, A. & Blundell, T. L.: Environment-specific amino acid substitution tables: tertiary templates and prediction of protein folds. *Protein Sci.* **1**, 216–226 (1992).

„Svoluji k zapůjčení této práce pro studijní účely a prosím, aby byla řádně vedena evidence vypůjčovateli.“

Jméno a příjmení S adresou	Číslo OP	Datum vypůjčení	Poznámka